Downloaded from orbit.dtu.dk on: May 05, 2023

DTU Library

Fed-Batch Process Modelling for Monitoring, Optimisation & Control. (incl. A Continuous Time Stochastic Modelling Framework)

Rasmussen, Jan Kamyno

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Rasmussen, J. K. (2008). *Fed-Batch Process Modelling for Monitoring, Optimisation & Control. (incl. A Continuous Time Stochastic Modelling Framework)*.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Databased and mechanistic modelling of fed-batch cultivation

Jan Kamyno Rasmussen

July, 2007

Computer Aided Process Engineering Center
Department of Chemical Engineering
Technical University of Denmark
and
Fermentation Pilot Plant
Novozymes A/S

Preface

This thesis is submitted to the Technical University of Denmark (DTU) in partial fulfillment of the requirements for the Ph.D. degree in Chemical Engineering. The work presented has been carried from September 2003 to August 2006 and has been funded by the Novozymes Bioprocess Academy which is a collaboration between Novozymes A/S and DTU. The work has been carried out at Computer Aided Process Engineering Center (CAPEC), Department of Chemical Engineering at DTU and at the Novozymes Fermentation Pilot Plant in Bagsværd, Denmark.

Throughout my work a number of people have provided their guidance and support, for which I am very grateful. First of all I would like to thank my supervisor Professor Sten Bay Jørgensen (CAPEC) and my co-supervisors Henrik Steen Jørgensen (Novozymes) and Henrik Madsen (Informatics and Mathematical Modeling, DTU) for their inspiration and advice.

I would like to thank the people at Novozymes whom I have worked with and my colleagues at the Department of Chemical Engineering.

I would like to thank the Novozymes Bioprocess Academy for financial funding of this project.

Kgs. Lyngby, July 2007

Jan Kamyno Rasmussen

Summary

The topic of this thesis is analysis of experimental data from the cultivation of *Aspergillus oryzae* at Novozymes A/S. The aim is to be able to develop predictive models which can be used in advanced control structures for the cultivation process. Introduction of model based control schemes yields possibilities for better control of the cultivation process and thereby achieving higher product yields.

To begin with historical data for cultivation of *Aspergillus oryzae* at the Novozymes Fermentation Pilot Plant has been studied and the data consistency has been investigated. The investigation indicates that carbon recoveries of more than 100% are achieved which indicates unaccuracies in one or more of the measurements.

Analyses of the seed cultivation and the batch part of the main cultivation have been carried out based on a grey-box stochastic modelling framework. It is shown that there is some correlation between the amount of biomass produced in the seed cultivation and the duration of the batch part of the main cultivation. This can prove useful for choosing a desired feeding strategy for the remainder of the main cultivation. During the batch part of the main cultivation the specific growth rate starts to decrease which can be explained by some limitation occuring at this point in the cultivation. It is observed that lowering the initial starch concentration in the batch medium increases the yield coefficient of carbon conversion into biomass. This means that less carbon dioxide is produced which indicates that low substrate concentrations support a more efficient metabolism.

The observed unaccuracies in data lead to the decision of carrying out an experiment with additional measurements. This experiment reveals that two metabolites, malate and mannitol, which have not been oberved before are formed. Malate is formed during periods of high glucose concentration which is thought to be a result of an overflow metabolism. During periods of insufficient supply of glucose the formed malate can substitute glucose. Mannitol formation has been observed during periods of oxygen limitation as a product of substrate level phosphorylation. The highest yields of enzyme product are achieved when the DOT level is above 6%. It has been observed that large pellets are transferred from the seed cultivation tank to the main cultivation tank. These are broken up to small fragments during the batch part and a balance between small and large particles exists during the fed-batch part. During oxygen limited conditions formation of large particles which are likely to be freely dispersed hyphae with long filaments is observed.

Further analysis of the experiment with additional measurements shows that the entire cultivation can be divided into distinct operating regions. Analysis of reaction rates during these operating regions leads to formation of a model for the cultivation process. Simulation shows that this model has some shortcomings as only one type of biomass exists. Introduction of an active and inactive fraction of the biomass improves the simulation results. This supports the hypothesis that a significant amount of the total biomass is inactive and therefore not contributing to the enzyme production.

A black-box modelling framework called Grid of Linear Models (GoLM) has been applied to model data from the industrial cultivation process of *Aspergillus oryzae*. This modelling framework divides the entire batch duration into a number of grid points and fits locally linear models to each of these grid points. Combination of these local models allows for approximation of the behaviour of the entire cultivation. The results show that in general good predictive capabilities can be obtained. A GoLM model has been fitted to the pilot plant data and a controlled cultivation has been carried using that model in combination with a Model Predictive Control (MPC) framework. During the controlled cultivation large disturbances in the air flow rate occured which makes it difficult to assess the performance of the controller.

Resumé på dansk

Emnet for denne tese er analyse af eksperimentelle data fra kultiveringen af *Aspergillus oryzae* på Novozymes A/S. Mlet er at blive i stand til at udvikle prædiktive modeller som kan bruges i avancerede reguleringsstrukturer til kultiveringsprocessen. Indførsel af modelbaseret regulering rummer muligheder for bedre regulering af kultiveringsprocessen og derved at opnå højere produktudbytter.

Til at begynde med er historiske data fra kultiveringen af *Aspergillus oryzae* på Novozymes A/S Fermentation Pilot Plant studeret og datas konsistens er undersgt. Undersøgelsen viser at carbon udbyttegrader på over 100% er opnået hvilket indikerer unøjagtigheder i en eller flere af målingerne.

Analyser af podekultiveringer og batch delen af hovedkultiveringen er blevet udført baseret på en grey-box stokastisk modelleringsmetode. Det er vist at der eksisterer en sammenhæng mellem mængden af biomasse produceret i podekultiveringen og længden af batch perioden i hovedkultiveringen. Dette kan vise sig nyttigt for at vælge en ønsket fødestrategi for resten af hovedkultiveringen. Under batch delen af hovedkultiveringen begynder den specifikke væksthastighed at aftage hvilket kan forklares ved at en begrænsning opstår på dette tidspunkt i kultiveringen. Det er observeret at en sænkning af den oprindelige mængde stivelse i batch mediet øger udbyttekoefficienten af omdannelse af carbon til biomasse. Dette betyder at mindre carbon dioxid dannes hvilket indikerer at lave substratkoncentrationer understøtter en mere effektiv metabolisme.

De observerede upræcisheder i data førte til en beslutning om at udføre et eksperiment med yderligere målinger. Dette eksperiment afslører at to metabolitter, malat og mannitol, som ikke er observeret før, dannes. Malat dannes under perioder med høj glukosekoncentration hvilket menes at være resultat af en overflow metabolisme. Under perioder med utilstrækkelig tilførsel af glukose kan den dannede malat substituere glukose. Mannitoldannelse er observeret under perioder med oxygenbægrænsning som et resultat af substratniveau phosphorylering. De højeste udbytter af enzymprodukt opnås nå DOT niveauet er over 6%. Det er observeret at store pellets overføres fra podekultiveringstanken til hovedkultiveringstanken. Disse slås i stykker til små fragmenter under batch delen og en balance mellem små og store partikler eksisterer under fed-batch delen. Under oxygenbegrænsede forhold observeres en dannelse af store partikler som menes at vre frit dispergerede hypher med lange filamenter.

Videre analyse af eksperimentet med yderligere målinger viser at hele kultiveringen kan opdeles i forskellige operationsområder. Analyse af reaktionshastighederne under disse operationsområder leder til dannelsen af en model for hele kultiveringsprocessen. Simuleringer viser at denne model har nogle mangler da kun en type biomasse eksisterer. Indførslen af en aktiv og inaktiv del af biomassen forbedrer simuleringsresultaterne. Dette understøtter hypotesen om at en signifikant del af den samlede biomasse er inaktiv og derfor ikke bidrager til enzymproduktionen.

En black-box modelleringsmetode kaldet Grid of Linear Models (GoLM) er anvendt

til at modellere data fra den industrielle kultiveringsproces af *Aspergillus oryzae*. Denne modelleringsmetode opdeler hele batch perioden i et antal gitterpunkter og tilpasser lokalt linere modeller til hver af disse gitterpunkter. Kombinationen af disse lokale modeller tillader approksimering af opførslen af hele kultiveringen. Resultaterne viser at der generelt kan opnås gode prædiktive egenskaber. En GoLM model er blevet tilpasset pilot plant data og en reguleret kultivering er blevet udført ved anvendelse af denne model i kombination med en Model Predictive Control (MPC) metode. Under den regulerede kultivering forekom der store forstyrrelser i lufttilførselshastigheden hvilket gør det vanskelligt at vurdere regulatorens præstation.

Contents

Appendices

1

Introduction

1.1 Motivation for modelling cultivation processes

Batch and fed-batch processes are widely used in chemical and biochemical industry for a large variety of processes. Cultivations are widely used in biochemical industry and these are most often carried out in fed-batch operations. The advantages of fed-batch operation of cultivation processes are many: Ability to control substrate concentrations, avoid strain instability, minimise problems of contamination etc. Common requirements for batch and fed-batch operation is that a certain product quantity of specified quality should be obtained within a defined time frame. These demands require that the process variability is kept within narrow limits. Reaching these objectives is most often not an easy task due to critical process variables not being observable as well as a lack of defined corrective actions in case of the process not following the desired trajectory. Methods and tools which can improve observability of critical process variables and ensure a more consistent production are therefore highly desired.

This work has been carried out in collaboration with Novozymes A/S which is currently the largest producer of enzymes in the world. Today enzymes are used for many industrially important processes and often substitute traditional chemical processes which reduces the environmental impact. The process studied in this work is cultivation of the filamentous fungus *Aspergillus oryzae* producing an α-amylase product called Fungamyl.

The current cultivation operation at Novozymes A/S is based on following predefined information given in the recipe for the specific product. Each recipe contains information on raw materials, initial conditions, allowable bounds on process variables, conditions for switching operating modes etc. Thus combining the recipe with the specifications for the particular equipment used one obtains the socalled batch operations model for the particular production. When deviations from the batch operations model occur action is taken by the process operators to bring the process back to the specified operating path. The corrective action depends heavily on the specific operator and his or her experience. It is highly desirable to achieve a more similar fault correction for comparable deviations. The lack of advanced control systems necessitates large bounds on the variable paths and results in a non-optimal utilisation of the production facilities.

The potential for improvement is illustrated in figure 1.1 which gives the measured enzyme activity trajectory for five industrial cultivations for the production of Fungamyl by the *Aspergillus oryzae* microorganism. All cultivations have been run

using the same recipe and some of them have had operator intervention because one or more variables have been outside the specified regions. The enzyme activity is a measure of the enzyme concentration of the broth and is the most important quality variable for this process.

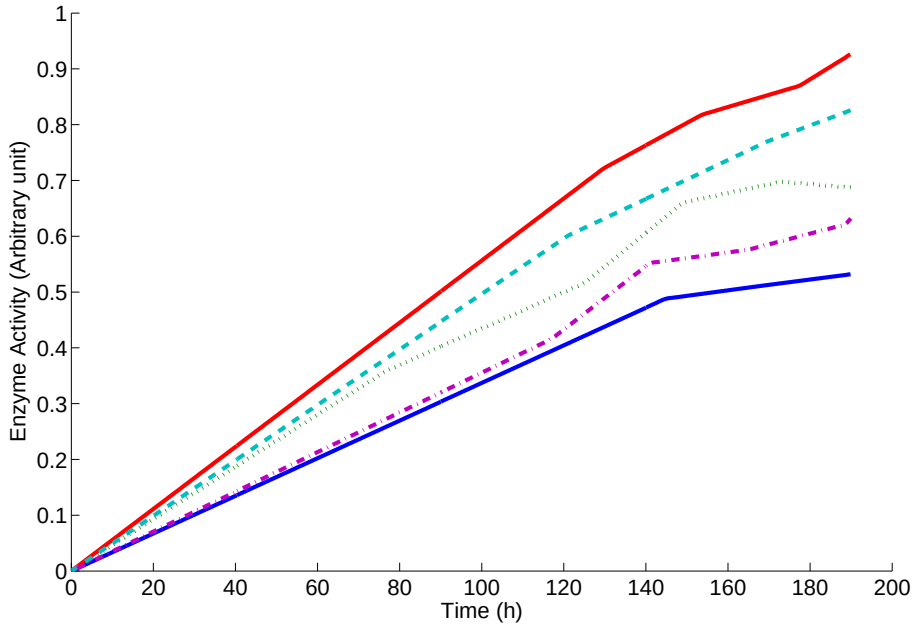


Figure 1.1. Example of enzyme activity trajectory for five industrial batches. The measured enzyme activity is plotted as a function of time. Interpolation between measured values has been used

Figure 1.1 shows that large variations in final enzyme activity occur even though the process has been run using identical recipes.

Reducing variability can be addressed by application of advanced multivariable control schemes. However such automated control systems rely on models which can be formulated mathematically. At the present level of limited understanding of the regulatory networks in microorganisms development of such models must be based on combining process knowledge and cultivation data. Being able to provide a predictive model of the cultivation process is the key for employing advanced control methods which have the potential to achieve more uniform variable trajectories and ensure consistently higher production yields.

1.2 Hypothesis

The key hypotheses of this work can be given in the following statements:

1. The data extracted from a cultivation at industrial scale is reproducible and contains sufficient information to develop predictive models of the evolution of the cultivation

2. The observed phenomena at specified operating conditions are repeated from one cultivation to the next and they occur at the same biological progress of the cultivation

3. The future goal of developing predictive models is: Based upon predictive models it is possible to achieve (much) more reproducible operation by using model predictive control

1.3 Contents

Chapter 2 gives a description of the modelling tools used in this work. Two major modelling frameworks have been used: Black-box modelling and Grey-box Stochastic Modelling. A short introduction to the physiology of the filamentous fungus studied is also given.

Chapter 3 provides information regarding the experimental setup and the measured variables in the data available from the Fermentation Pilot Plant at Novozymes A/S. Treatment of data and an assesment of their quality is also given in this chapter.

Chapter 4 analyses the seed cultivation in the Novozymes process. The outcome of the seed cultivation is crucial to a succesful operation of the subsequent main cultivation

Chapter 5 analyses the main cultivation where the major amount of the desired product is formed.

Chapter 6 reports a detailed experiment carried out at the Fermentation Pilot Plant. This experiment has been carried out with extensive sampling and sample analysis in order to obtain more information on the prevailing mechanisms in this cultivation process.

Chapter 7 analyses the results obtained in chapter 6 in more detail. Two models which cover the entire cultivation are proposed.

Chapter 8 provides an investigation of the Grid of Linear Models framework on data sets obtained from an industrial plant.

Chapter 9 provides a suggestion for a controller framework based on the Grid of Linear Models framework. This controller framework has been implemented in the Fermentaton Pilot Plant and a controlled cultivation has been carried out.

Chapter 10 concludes this thesis and presents ideas for future work.

1.4 Publications

1.4.1 Conference proceedings

Jan Kamyno Rasmussen, H. Madsen, and Sten Bay Jørgensen: Grey-box stochastic modelling of industrial fed-batch cultivation, Computer Aided Chemical Engineering, 21 421-426, 2006

Jan Kamyno Rasmussen and Sten Bay Jørgensen: Modelling and control of industrial fermentation, World Congress of Chemical Engineering, Glasgow, 2005

Jan Kamyno Rasmussen, H. Madsen, and Sten Bay Jørgensen: Modelling for Control of Industrial Fermentation, ESCAPE 15, Barcelona, Spain, 29 May - 1 June, 1351-1356, 2005

Jan Kamyno Rasmussen and Sten Bay Jørgensen: Modelling and Control of Industrial Fermentation, AIChE Spring Meeting 2005, Atlanta, USA, 10-14 April, 2005

Jan Kamyno Rasmussen and Sten Bay Jørgensen: Modeling and Control of Industrial Fermentation, AIChE Annual Meeting 2004, Austin, USA, 7-12 November, 2004

Jan Kamyno Rasmussen and Sten Bay Jørgensen: Experimental Investigation of Datadriven Modelling and Control of Fed-Batch Cultivation, IFAC 2004, Prague, Czech Republic, 4 - 8 July, 2004

Jan Kamyno Rasmussen, H. Madsen, and Sten Bay Jørgensen: Modeling and control of industrial fermentation, Nordic Process Control Workshop 12, 2004

Jan Kamyno Rasmussen, H. Madsen, and Sten Bay Jørgensen: Modeling and Control of Industrial Fermentation, BatchPro Symposium, Poros, Greece, 6-9 June, 335-342, 2004

Jan Kamyno Rasmussen and Sten Bay Jørgensen: Modeling and Control of Industrial Fermentation, Dansk Kemiingenirkonference, DK2, DTU, 24-26 May, 2004

1.4.2 Conference presentations

Jan Kamyno Rasmussen, Sten Bay Jørgensen, and H. Madsen: Stochastic Grey-box Modelling of Industrial Fed-batch Fermentation, Dansk Kemiingeniør Konference - DK2 - 31. maj - 2. juni 2006, 164-165, 2006

Jan Kamyno Rasmussen, H. Madsen, and Sten Bay Jørgensen: Stochastic Grey-box Modelling of Industrial Fermentation, AIChE Annual Meeting 2005, Cincinnati, USA, Oct. 29-Nov. 4, 2005

1.5 Confidentiality

Some of the data used in this work are regarded confidential by Novozymes A/S and can therefore not be published unedited. In order to minimise the loss in understanding and explanation of phenomena it has been decided to use a scaling factor for all data related to the enzyme yield, ie. numbers, tables and figures. Also the actual numbers for oxygen uptake and carbon dioxide evolution are not given.

Part I

Background

2

Background methods and knowledge

2.1 Introduction

This chapter provides background information on the methodologies and concepts used in this work. An overview of the the theoretical frameworks used is given while the theory is covered in detail in the literature. The scope of this thesis is to model industrial fed-batch cultivations of *Aspergillus oryzae* and the contents reflect this scope.

Section 2.2 gives an overview of the main types of modelling methodologies. Section 2.3 and 2.4 provide an introduction to the Grey-box stochastic modelling framework and to the Grid of Linear Models framework respectively. These two are the major modelling frameworks used in this work. Section 2.5 provides key aspects of the physiology of the micro-organism studied in this thesis and finally a conclusion is given in section 2.6.

2.2 Modelling methodologies

Developing mathematical models can be approached in numerous ways. Traditionally mathematical conservation law based models mostly have been formulated by designing laboratory scale experiments and investigating specific static or dynamic properties of the system. This is often associated with extensive and laborious efforts. This is known as traditional first principles engineering modelling or white-box modelling. In recent years more focus has been directed towards the data-driven or black-box modelling approach. In this approach models are formulated entirely based on input/output process data connected to the process. This approach partially circumvents the need for a thorough understanding of the system and allows for relatively fast model development compared to the first principles approach.

The choice of modelling method depends heavily on data availability and on the intended application of the model. Models can be developed for process design, monitoring, control, optimisation, etc. Each of the applications have different requirements regarding model accuracy, level of detail etc.

The major types of modelling methodologies are described in the following.

- **White-box modelling**
  White box or first principles engineering modelling is based on fundamental knowledge combined with information obtained from experimental data if

necessary. Usually such data are obtained in the laboratory under well determined conditions and the experiments are designed to investigate specific phenomena. These models generally have a very good physiological interpretation and are well suited for designing further experiments or implementation in local controllers. However the obtained models are most often too simple to apply in industrial cultivations.

- **Black-box modelling**
  Black-box modelling is mainly data-driven. Very limited knowledge of the process is needed and the models developed therefore usually also provide very limited possibilities for physiological interpretation. Data can be taken from large scale production plants but one has to be very cautious concerning the quality of such data. Data collected in industrial plants are traditionally not intended for modelling but are used by the operators to monitor the process and identify major faults. The data can therefore be very noisy and the sensors are not necessarily calibrated accurately and also variations in calibration from batch to batch can occur due to the hostile environment during sterilisation. However provided the data is of satisfactory quality robust models can be obtained with little effort. Examples of data-driven methodologies are: Principal Component Analysis (PCA) (MacGregor and Kourti (1995), Gregersen (2003)), Partial Least Squares (PLS) (Kourti *et al.* (1995), MacGregor and Kourti (1995), Gregersen (1999)), Neural networks and Grid of Linear Models (GoLM) (Bonné (2005)).

- **Grey-box modelling**
  As the name suggests grey-box modelling is a combination of white-box and black-box modelling. A first principles engineering model is constructed and experimental data are used to identify missing phenomena in the model and evaluate its quality. Different grey-box modelling approaches have been used and in this work the grey-box stochastic modelling framework suggested in Kristensen *et al.* (2004) has been employed. This methodology is based on the use of stochastic differential equations and the methodology is able to provide indications on how to improve the model structure.

In this work both the black-box and grey-box modelling approaches have been used, hence the specific methods are briefly introduced in the following sections.

2.3 Grid of Linear Models framework

The black-box modelling framework proposed by Bonné and Jørgensen (2003) is used in this work. The framework is called "Grid of Linear Models" (GoLM) and the approach is purely data-driven which means that only limited prior knowledge of the process is required. The basic assumptions of the mehodology is that the non-stationary and non-linear dynamic behaviour of batch processes can be approximated with sets of interdependent linear models (Bonné (2005)). In practice this is achieved by subdividing the time span of the entire process into a grid where each time interval contains a multivariable Linear Time Invariant (LTI) model. The

methodology is especially suited for modelling repeated batch or other periodic processes. The distance between the grid points can either be constant or dependent on the current condition or operating region of the process, thus reflecting the process dynamics.

This framework has been implemented as a toolbox for Matlab. The current version of the toolbox supports two different linear time series models: Auto Regressive with eXogenuos inputs (ARX) and Auto Regressive Moving Average with eXogenuos inputs (ARMAX). The identified time series models are represented as state space (SS) models for straightforward implementation in a model predictive control framework aimed specifically at control of batch processes (Bonné and Jørgensen (2004)).

2.3.1 Model structure

The following time series have been defined:

- Input variables $u_t \in \mathbb{R}^{n_u(t)}$
- Output variables $y_t \in \mathbb{R}^{n_y(t)}$
- Disturbance variables $w_t \in \mathbb{R}^{n_y(t)}$

and corresponding to the input and output variables their reference trajectories are, $\bar{u}_t \in \mathbb{R}^{n_u(t)}$ and $\bar{y}_t \in \mathbb{R}^{n_y(t)}$ respectively. The ARX model parameterisation is applied to describe the output deviation $\bar{y}_t - y_t$ at sample time t as a weighted sum of the past $n_b(t)$ input and $n_a(t)$ output deviations formulated as:

$$\begin{aligned} \bar{y}_t - y_t = &- a_{t,t-1}\big(\bar{y}_{t-1} - y_{t-1}\big) - \dots \\ &- a_{t,t-n_a(t)}\big(\bar{y}_{t-n_a(t)} - y_{t-n_a(t)}\big) \\ &+ b_{t,t-1}\big(\bar{u}_{t-1} - u_{t-1}\big) + \dots \\ &+ b_{t,t-n_b(t)}\big(\bar{u}_{t-n_b(t)} - u_{t-n_b(t)}\big) \\ &+ w_t \end{aligned} \tag{2.1}$$

Here $n_a(t)$ and $n_b(t) \in [1, \dots, t]$ define the model orders of the local ARX model at each grid point, while $a_{i,j} \in \mathbb{R}^{n_y(i),n_y(j)}$ and $b_{i,j} \in \mathbb{R}^{n_y(i),n_u(j)}$ are the parameters of the local ARX model.

Having an operation with N sampling points, the input $\boldsymbol{u}$, output $\boldsymbol{y}$, one sample time shifted output $\boldsymbol{y}^0$, and disturbance $\boldsymbol{w}$ profiles are defined as:

$$\begin{aligned} \boldsymbol{u} &= [u_0^T u_1^T \dots u_{N-1}^T]^T \\ \boldsymbol{y} &= [y_1^T y_2^T \dots y_N^T]^T \\ \boldsymbol{y}^0 &= [y_0^T y_1^T \dots y_{N-1}^T]^T \\ \boldsymbol{w} &= [w_1^T w_2^T \dots w_N^T]^T \end{aligned} \tag{2.2}$$

The ARX model can then be formulated in matrix representation:

$$\bar{\boldsymbol{y}} - \boldsymbol{y} = -\boldsymbol{A}(\bar{\boldsymbol{y}}^0 - \boldsymbol{y}^0) + \boldsymbol{B}(\bar{\boldsymbol{u}} - \ \boldsymbol{u}) + \boldsymbol{w} \tag{2.3}$$

where $\boldsymbol{A}$,$\boldsymbol{B}$ are structured lower block triangular matrices.

2.3.2 Batch to Batch Modeling

The disturbance profile $\boldsymbol{w}$ is composed of contributions from a number of sources which can be subdivided into repeated disturbances *e.g.* recipe/input bias, model bias and erroneous readings, and random disturbances *e.g.* process upsets with no correlation between subsequent batches.

By adopting the batch-to-batch approach from Iterative Learning Control Bonné and Jørgensen (2003) showed how it is possible to model the disturbance profile as a random walk with respect to the batch index k:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{v}_k \tag{2.4}$$

where $\boldsymbol{v} = [v_1^T v_2^T \dots v_N^T]^T \in \mathbb{R}^{n_y(t)\cdot N}$ represents the part of the disturbance sequence which is not repeated. This is assumed to be zero-mean, independent and identically distributed.

The difference between two successive operations then becomes:

$$\begin{aligned} \Delta\boldsymbol{y}_k &= \boldsymbol{y}_k - \boldsymbol{y}_{k-1} \\ &= \boldsymbol{A}(\boldsymbol{y}_k^0 - \boldsymbol{y}_{k-1}^0) - \boldsymbol{B}(\boldsymbol{u}_k - \boldsymbol{u}_{k-1}) \\ &\quad + \boldsymbol{w}_k - \boldsymbol{w}_{k-1} \\ &= \boldsymbol{A}\Delta\boldsymbol{y}_k^0 - \boldsymbol{B}\Delta\boldsymbol{u}_k + \boldsymbol{v}_k \end{aligned} \tag{2.5}$$

This is a model with an ARX structure, which models dynamics from batch to batch assuming the same reference profile with nonrepeated disturbances. ($\bar{\boldsymbol{y}}$,$\bar{\boldsymbol{u}}$) and of repeated disturbances.

2.3.3 Model Identification

For the estimation of the model parameters and model orders, a Least Squares (LS) methodology can not be used directly due to the many local models. However an interrelation between models in neighbouring grid points can be expected, especially when the time difference between the grid points is shorter than the time constants of the relevant process dynamics. Using Tikhonov regularisation the model parameters can be estimated by solving the extended LS problem:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{\boldsymbol{\Lambda}} &= \arg\min_{\boldsymbol{\theta}} \big[(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\theta}) \\ &\quad + (\boldsymbol{\Lambda}\boldsymbol{L}\boldsymbol{\theta})^T(\boldsymbol{\Lambda}\boldsymbol{L}\boldsymbol{\theta})\big] \\ &= (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} + \boldsymbol{L}^{\boldsymbol{T}}\boldsymbol{\Lambda}^{\boldsymbol{2}}\boldsymbol{L})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} \end{aligned} \tag{2.6}$$

where the penalty $\boldsymbol{\Lambda}\boldsymbol{L}\boldsymbol{\theta}$ is a column vector of weighted differences between parameters in neighbouring grid point models. $\boldsymbol{L}$ being a structured penalty matrix, maps the parameter vector $\boldsymbol{\theta}$ into the appropriate parameter difference where the diagonal regularisation matrix $\boldsymbol{\Lambda}$ finally weights these differences. The interdependency between the grid point models is determined by the choice of structure of the penalty matrix $\boldsymbol{L}$ and the values in the regularisation matrix, $\boldsymbol{\Lambda}$, as outlined in Bonné and Jørgensen (2003).

The model orders and regularisation weights can be estimated based on minimisation of the mean square prediction error obtained from using the proposed model on a validation data set. Different model properties can be obtained depending on the choice of prediction horizon for the prediction error *e.g.* one step ahead or pure simulation for the full batch length.

2.4 Grey-box stochastic modelling framework

The grey-box framework proposed in Kristensen *et al.* (2004) has been used in this work. The approach given in this framework is to combine first principle engineering models with operational data to produce predictive models suited for control purposes. The method is called grey-box stochastic modeling and combines a set of stochastic differential equations describing the dynamics of the system in continous time with a set of discrete time measurements. An important advantage using this approach compared to using deterministic models is that that stochastic models can account for random variations in data and thereby provide a sound basis for hypothesis testing related to model validity based on available data. The stochastic differential equations include both measurement and process noise terms which render it possible to distinguish noise from unmodelled dynamics.

One of the key ideas behind the the grey-box stochastic modelling framework is to use all prior information for formulation of an initial first principles engineering model. Unknown parameters for the initial model are then estimated from experimental data and an analysis of validation data residuals is carried out to evaluate the properties of the resulting model. On this basis the next step in the modeling cycle can be carried out. This step is the model falsification or unfalsification which determines if the model is sufficiently accurate to serve its intended purpose. If the model is unfalsified the model development is completed. In case of falsification the modeling cycle must be repeated either by reformulating the initial model or by by redesigning the experiment. In case of a falsified model different statistical tests can be used to provide indications of which parts of the model that are deficient. Nonparametric modelling is applied to estimate which functional relationships are needed to improve the model.

A summary of the modelling cycle is given in figure 2.1.

2.4.1 Model structure and estimation

This section describes briefly the estimation method, which is a maximum likelihood method for estimating parameters in continuous-discrete time stochastic state space models, where the time evolution is described by Stochastic Differential Equations (SDE's). A detailed outline is given in Kristensen *et al.* (2004) and Kristensen (2003).

In the general case, the continuous-discrete stochastic state space model is a model that consists of a set of nonlinear, partially observed Itô Stochastic Differential Equations (SDE's) with measurement noise, i.e.:

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{W}_t \tag{2.7}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k \tag{2.8}$$

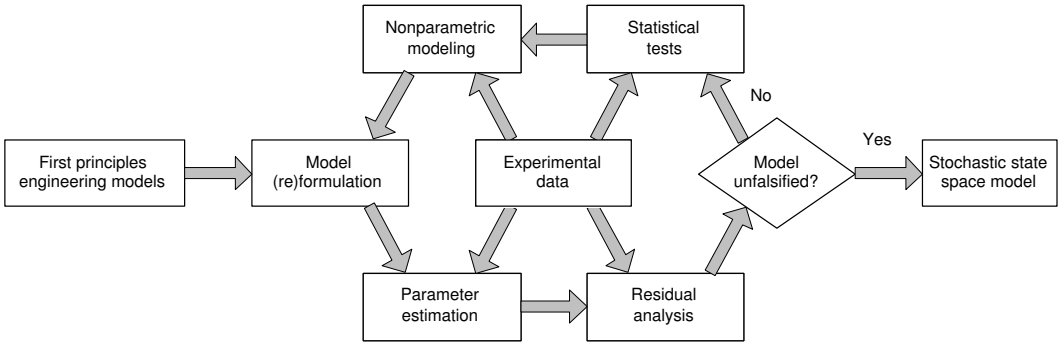


Figure 2.1. The grey-box stochastic modelling cycle proposed in Kristensen *et al.* (2004). The modelling cycle is initiated by combining an initial first principles engineering model with experimental data. The modelling cycle is continued until an unfalsified model is obtained

where $t \in \mathbb{R}$ is the time variable; $\boldsymbol{x}_t \in \mathcal{X} \subset \mathbb{R}^n$ is a vector of state variables; $\boldsymbol{u}_t \in \mathcal{U} \subset \mathbb{R}^m$ is a vector of input variables; $\boldsymbol{y}_k \in \mathcal{Y} \subset \mathbb{R}^l$ is a vector of output variables; $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a vector of (possibly unknown) parameters; $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$, $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$ and $\boldsymbol{h}(\cdot) \in \mathbb{R}^l$ are nonlinear functions; $\{\boldsymbol{W}_t\}$ is an n-dimensional standard Wiener process and $\{\boldsymbol{e}_k\}$ is an l-dimensional Gaussian white noise process with $\boldsymbol{e}_k \in N(\boldsymbol{0}, \boldsymbol{S}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}))$. The Wiener process and $\{\boldsymbol{e}_k\}$ are assumed independent. Note that the diffusion term in (2.7) is assumed to be independent of the state variables, because this renders parameter estimation more feasible.

2.4.2 Parameter estimation

The solution to (2.7) is a Markov process and an estimation scheme based on probabilistic methods can therefore be applied to estimate the unknown parameters of the model in (2.7)-(2.8), e.g. *maximum likelihood* (ML) or *maximum a posteriori* (MAP), where the latter can be applied if prior information about the parameters is available. Let:

$$\mathbf{Y} = [\mathcal{Y}_{N_1}^1, \mathcal{Y}_{N_2}^2, \ldots, \mathcal{Y}_{N_i}^i, \ldots, \mathcal{Y}_{N_S}^S] \tag{2.9}$$

be a set of S stochastically independent sequences of consecutive measurements, where:

$$\mathcal{Y}_{N_i}^i = [\boldsymbol{y}_{N_i}^i, \ldots, \boldsymbol{y}_k^i, \ldots, \boldsymbol{y}_1^i, \boldsymbol{y}_0^i] \tag{2.10}$$

and let $p(\boldsymbol{\theta})$ be a prior probability density function for the parameters. In the general case, point estimates of the parameters in (2.7)-(2.8) can then be found as the parameters $\boldsymbol{\theta}$ that maximize the posterior probability density function:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left(\prod_{i=1}^{S} p(\mathcal{Y}_{N_i}^i|\boldsymbol{\theta}) \right) p(\boldsymbol{\theta}) \tag{2.11}$$

or equivalently:

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left(\prod_{i=1}^{S} \left(\prod_{k=1}^{N_i} p(\boldsymbol{y}_k^i|\mathcal{Y}_{k-1}^i, \boldsymbol{\theta}) \right) p(\boldsymbol{y}_0^i|\boldsymbol{\theta}) \right) p(\boldsymbol{\theta}) \tag{2.12}$$

where Baye's rule $P(A \cap B) = P(A|B)P(B)$ has been applied successively to form products of conditional probability density functions. By neglecting the prior densities the likelihood function is obtained. The posterior density function and the likelihood function are thus a product of conditional probability densities.

Nielsen *et al.* (2000) review the state of the art of parameter estimation in discretely observed Itô SDE's, which shows that for models of the type in (2.7)-(2.8) only methods based on approximate nonlinear filters provide a computationally feasible solution to the problem. However, since the diffusion term in (2.7) has been assumed to be independent of the state variables, a much simpler alternative can be applied here.

More specifically, since the SDE's in (2.7) are driven by a Wiener process, and since increments of a Wiener process are Gaussian, it is reasonable to assume, under some regularity conditions, that the conditional probability density functions can be well approximated by Gaussian densities, which means that a method based on the extended Kalman filter (EKF) can be applied. The Gaussian density is completely characterised by its mean and covariance, hence by introducing:

$$\hat{\boldsymbol{y}}^i_{k|k-1} = E\{\boldsymbol{y}^i_k|\mathcal{Y}^i_{k-1}, \boldsymbol{\theta}\} \tag{2.13}$$

$$\boldsymbol{R}^i_{k|k-1} = V\{\boldsymbol{y}^i_k|\mathcal{Y}^i_{k-1}, \boldsymbol{\theta}\} \tag{2.14}$$

$$\boldsymbol{\epsilon}^i_k = \boldsymbol{y}^i_k - \hat{\boldsymbol{y}}^i_{k|k-1} \tag{2.15}$$

and by further assuming that the prior probability density function for the parameters is Gaussian as well and therefore also introducing:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} \tag{2.16}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = V\{\boldsymbol{\theta}\} \tag{2.17}$$

$$\boldsymbol{\epsilon}_{\boldsymbol{\theta}} = \boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}} \tag{2.18}$$

the posterior probability density function becomes:

$$\begin{aligned} &p(\boldsymbol{\theta}|\mathbf{Y}) \propto \\ &\left(\prod_{i=1}^{S}\left(\prod_{k=1}^{N_i}\frac{\exp\left(-\frac{1}{2}(\boldsymbol{\epsilon}^i_k)^T(\boldsymbol{R}^i_{k|k-1})^{-1}\boldsymbol{\epsilon}^i_k\right)}{\sqrt{\det\left(\boldsymbol{R}^i_{k|k-1}\right)}\left(\sqrt{2\pi}\right)^l}\right)\right. \\ &\left.\times p(\boldsymbol{y}^i_0|\boldsymbol{\theta})\right) \times \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}^T_{\boldsymbol{\theta}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\theta}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}\right)}{\sqrt{\det\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)}\left(\sqrt{2\pi}\right)^p} \end{aligned} \tag{2.19}$$

The parameter estimates can now be determined by further conditioning on the initial conditions:

$$\mathbf{y_0} = [\boldsymbol{y}^1_0, \boldsymbol{y}^2_0, \ldots, \boldsymbol{y}^i_0, \ldots, \boldsymbol{y}^S_0] \tag{2.20}$$

and applying nonlinear optimisation to find the minimum of the negative logarithm of the resulting posterior probability density function, i.e.:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta}\left\{-\ln\left(p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{y_0})\right)\right\} \tag{2.21}$$

In the general case the estimation method implied by (2.21) is a maximum a posteriori method, but, if no prior information about the parameters is available, it

reduces to a maximum likelihood method. In either case, the innovations $\boldsymbol{\epsilon}_k^i$ and their covariances $\boldsymbol{R}_{k|k-1}^i$ can be computed recursively by means of an EKF for each set of parameters $\boldsymbol{\theta}$ in the optimisation, see Kristensen *et al.* (2004) for further details.

2.4.3 Uncertainty of parameter estimates

A sound statistical parameter estimation scheme must include an assessment of the uncertainty of the estimates. An estimate of the uncertainty of the parameter estimates can be obtained by using the fact that by the central limit theorem the estimator in (2.21) is asymptotically Gaussian with mean $\boldsymbol{\theta}$ and covariance matrix:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{H}^{-1} \tag{2.22}$$

where the matrix $\boldsymbol{H}$ is given by the elements:

$$h_{ij} = -E\left\{\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{y_0})\right)\right\} \approx -\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln\left(p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{y_0})\right)\right)\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

for $i,j = 1,\ldots,p$, and where an approximation to $\boldsymbol{H}$ can be obtained at the minimum of the objective function. A measure of the uncertainty of the individual estimates can be obtained by decomposing the covariance matrix:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}\boldsymbol{R}\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}} \tag{2.23}$$

into $\boldsymbol{\sigma}_{\hat{\boldsymbol{\theta}}}$, which is a diagonal matrix of the standard deviations of the parameter estimates, and $\boldsymbol{R}$, which is the corresponding correlation matrix.

The uncertainty information thus obtained can subsequently be applied to perform tests of various hypotheses, e.g. to determine the significance of the individual parameters through t-tests.

2.4.4 Software Implementation

The parameter estimation scheme has been implemented in a computer program called **CTSM**. This program is available for both Linux, Solaris and Windows platforms (The software is available from `www.capec.kt.dtu.dk`). A graphical user interface (GUI) allows the user to specify the model equations and how each parameters should be estimated. After specifying which data sets to use, the program then determines the parameter estimates and computes the uncertainty information. The program also provides features to deal with occasional outliers and missing observations, and on Solaris systems the program supports shared memory parallelisation to alleviate the extensive computational load often associated with estimation of parameters in continuous-discrete stochastic state space models.

2.5 Filamentous fungal physiology

This section focuses on introducing some basic information regarding the physiology of the microorganism studied in this work. Each microorganism has its own characteristics which are important to understand in order to design experiments and

interpret experimental results properly. The micro-organism studied in this work is the filamentous fungus *Aspergillus oryzae* and the following discussion aims at giving a basic understanding of important concepts related to this micro-organism.

2.5.1 Fungal morphology

Morphology deals with the form and structure of the micro-organism. The morphology can be classified into microscopic and macroscopic morphology (Nielsen and Villadsen (1992)). The microscopic morphology deals with the characterisation of single hyphal elements. This includes properties like hyphal length, number of tips, diameter of tips etc. The macroscopic morphology deals with characterisation of a collection of hyphal elements. The major macroscopical divisiones are freely dispersed hyphae and pellet.

Filamentous fungi grow in a very different manner compared to unicellular micro-organisms. The cells in filamentous fungi are connected in structures called hyphae. All cells in the hyphal structure may contribute to the growth process by producing necessary constituents but extension of hyphae can only occur at the tips. The excretion of extracellular proteins eg. enzymes also occurs at the tips and is associated to hyphal growth. The number of tips per unit (length or mass) of hyphae is a characteristic morphological variable Nielsen *et al.* (2003). As the hyphae grow the inner (and oldest) parts of the hyphae can become evacuated of cell material and thus become inactive.

Freely dispersed growth is characterised by no particular interconnection between the hyphal structures. Pellets on the other hand are characterised by hyphae gathering in large agglomerates. When pellets are formed the interior of the pellet might suffer heavily from substrate and oxygen limitation which reduces the enzyme production rate and eventually leads to the death of cells. It is often observed that the inner part of the pellet is in fact dead.

Pellet formation of *Aspergillus oryzae* has been studied by Carlsen *et al.* (1996b) where it was found that the formation of free hyphae or pellets is closely related to the pH level of the medium. The macroscopical morphology depends on a number of other factors as well, among others are stirring speed, trace metals, type of strain etc. The influence of different factors on morphology is illustrated in figure 2.2.

2.5.1.1 Pellet structure

The pellet structure is often classified into 3 different types (Metz and Kossen (1977)):

- Fluffy loose pellets: The core is quite dense and the outer regions are much looser. The surface is hairy.
- Compact smooth pellets: The entire pellet is dense and the surface is smooth.
- Hollow smooth pellets: Pellets are dense and smooth on the surface but the inside is hollow due to cell lysis (rupture and destruction of cells). This type occurs if the pellet center becomes limited in one or more of the critical substrates.

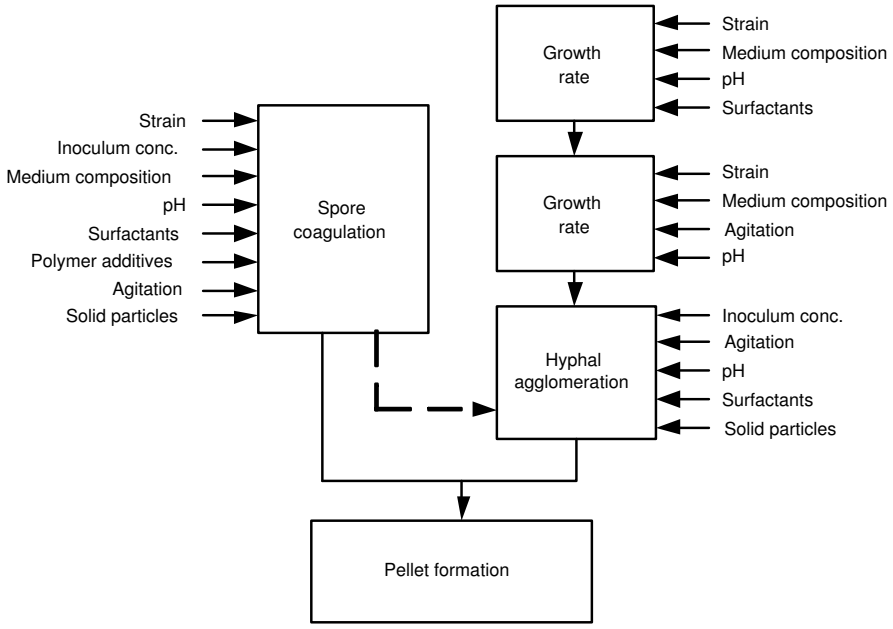


Figure 2.2. Factors influencing filamentous fungal morphology. From Metz and Kossen (1977)

Different methods have been developed to characterise morphology in an automated way. Eg. Reichl *et al.* (1992) used automated image analysis to characterise morphology and divide the biomass into pellets and free hyphae. Poulsen (2005) developed a 'Macro-morhology Profiling System' (MPS) for quantification of macromorphology of submerged cultures of filamentous microorganisms. Common for these methods are that they are difficult to implement and run on industrial scale cultivation tanks.

2.5.1.2 Limitations in growth and production rate

One or more factors will always limit the growth and/or production rate of the micro-organism. If all extracellular conditions are optimal the limitation will be due to bottlenecks in the metabolic pathways in the cell. Industrial scale cultivations will most often be affected by limitations in extracellular conditions, e.g. oxygen limitation or substrate limitation. Oxygen limitation occurs if regions of the cells are subject to an insufficient oxygen concentration in the cultivation broth surrounding them. The oxygen transfer rate to the different parts of the micro-organism is affected by a number of factors, e.g. aeration rate, gas bubble size and broth viscosity. The macroscopic morphology has a very large impact on the properties of the fermentation broth. Zangirolami (1998) found that lack of oxygen also leads to a drastic change in the morphology. Under these circumstances it was observed that long filaments were grown which further increase viscosity and impairs oxygen transfer even further. Formation of pellets also leads to increased viscosity and reduced mass transfer coefficients which has a big impact on the growth rate of the fungus (Agger (1999)). High viscocity of the medium also leads to zones of poor mixing causing concentration gradients. For industrial cultivation tanks it has

been observed that large gradients can occur (Li *et al.* (2002)). This can lead to some cells being subject to high glucose concentrations which inhibit enzyme formation. The overall performance of the process might therefore be seriously affected by insufficient mixing.

Limitations in substrate level can lead to reduced biomass growth and in severe cases even to cell death. Low substrate concentration however favours production of enzyme and in industrial applications the substrate level is kept low during in the major part of the cultivation in order to stimulate production of enzyme in preference to biomass.

Under certain conditions as starvation or very high shear stress the hyphae can fragment. Vacuated regions have a higher tendency to break but also actively growing biomass can fragment. This phenomenon is called cell lysis and the cell material is simply emptied out into the fermentation broth.

2.5.2 Overview of published models

Substantial effort has been made to mathematically model processes involving microorganisms. Most attention has been given to simple micro-organims like bacteria and yeast but also models for more complicated micro-organisms such as filamentous fungi have been developed.

However, relatively little work has been concerned with applications working directly on industrial data. This is often connected to confidentiality issues and lack of possibilities of proper data collection and treatment.

Existing models describing the filamentous fungal cultivation can be found in the literature. Fungal growth is very different from bacterial growth because fungi are characterized by growing hyphae with properties depending on age and growth conditions. This makes compartment models especially attractive for modelling fungal growth because each type of fungal region and the transition between regions can be described mathematically.

The strain studied in this work is a high yielding strain (with respect to enzyme formation) while the strains described in literature are usually not optimised for enzyme production. The yield coefficients and other parameters found in the literature are therefore not directly comparable to the results obtained in this work.

2.5.2.1 Structured and unstructured models

The models developed for filamentous fungal cultivations can be classified into two main groups: Structured and unstructured biomass models. Unstructured models describe the biomass by only one variable, usually the biomass concentration or total biomass weight. Structured models consider the structure of the biomass and therefore contain more variables for the biomass description. Structured models have varying degrees of complexity. The simplest ones divide the biomass into different compartments whereas the more complicated ones consider internal fluxes and activities of the individual enzymes. Unstructured models are usually based on empirical relationships and are often adequate to explain growth and product formation. For description of rapidly changing growth conditions models with some kind of structure are usually necessary (Nielsen *et al.* (2003)). Especially for fila-

mentous fungal fermentations the morphology is an important factor and this can be described by a segregated model. Segregated models consider different morhological forms and hereby distinguish cells with different internal compositions from each other (Nielsen *et al.* (2003)).

Such morhphologically structured models have been proposed for *Aspergillus awamori* (Megee *et al.* (1970)), *Penicillium chrysogenum* (Nestaas and Wang (1983); Paul and Thomas (1996)) and *Aspergillus oryzae* (Brown and Fitzpatrick (1979)). As mentioned earlier in this chapter quantifying hyphal differentiation is technically challenging but in the work of Paul and Thomas (1996) image analysis was used to estimate parameters in a structured model for *Penicillium chrysogenum*. Extensive studies have been carrried out on the α-amylase production of *Aspergillus oryzae* in Carlsen *et al.* (1996a) and Spohr *et al.* (1998a). Spohr *et al.* (1998b) used a flow-through cell combined with image analysis to develop a model for the growth kinetics of hyphae of *Aspergillus oryzae*. This model has been used as a basis to develop a morphologically structured model of *Aspergillus oryzae* which contains a description of biomass and product formation as well as properties of the hyphae. A drawback of this model is that it contains a large number of parameters and especially the ones describing hyphal growth are not easy to estimate. Also the model operates with three different types of biomass which are not identifiable using industrial on-line measurements.

2.5.2.2 Modelling of pellet formation

Pellet formation of filamentous fungi has been studied for a number of years due to the large impact on industrial fermentations. Pirt (1966) studied formation of pellets and recognised that the inner parts of the pellets may very well become limited in one or more substrates leading to extensive metabolic differences between outer and inner cells.

A critical pellet diameter can be defined (Carlsen (1994)) which is the largest diamater the pellet can have without being subject to diffusion limitations. The critical pellet diameter depends heavily on the limited substrate. When glucose is limiting the critical diameter is typically 2-3 times lower than when oxygen is limiting. Diffusion limitation during periods of high substrate and oxygen concentration (eg. the batch period of the cultivation studied) can result in non-exponential biomass grow (Poulsen (2005)). Provided that satisfactory biomass measurements or estimates are available the shift from exponential to non-exponential growth can be used to indicate at which point in time the limitation sets in.

2.6 Conclusion

This chapter has introduced the methodologies and concepts used in this thesis. The two major modelling frameworks, namely Grid of Linear Models (GoLM) and Stochastic Grey-box Modelling have been presented. A review of the physiology of *Aspergillus oryzae* is given as well as a literature review of existing mathematical models for cultivation processes with filamentous organisms.

3

Experimental plant description

3.1 Introduction

The purpose of this chapter is to introduce the experimental setup which provides the cultivation data for this thesis. Hence the description is aimed at providing insight into the origin and quality of the process data used in this work. Understanding of how data have been generated and assessing how reliable they are is very important for the subsequent analysis.

Data is generated in two fundamentally different ways: They are either on-line data or off-line data. On-line data is typically generated by sensors located on or very close to the cultivation tank making continuous measurements of some property. Off-line data are typically generated by taking cultivation broth out of the cultivation tank for subsequent analysis in the laboratory. Different factors need to be considered in each case when the data is to be evaluated. Regarding online sensors both the location and measurement principle is of great importance to the quality of the data. The sensor location must be known in order to know whether the property measured is representative for the whole tank or not. Regarding offline measurements both measurement principles and the uncertainty of the experimental procedure is important to know before further analysis is carried out. The measurement uncertainty is often much larger for off-line measurements than for on-line measurements.

The importance of understanding how data is generated and understanding whether it provides a representative picture of the process is often underestimated. Without this understanding process data can not be used to draw definite conclusions.

This chapter first provides an introduction to the process studied which is given in section 3.2. This is followed by section 3.3 which introduces the variables available from the process control system in the pilot plant. Section 3.4 then describes how the variables have been treated prior to further analysis. Section 3.5 uses material balances to assess the quality of the available data. Finally conclusions are given in section 3.6.

3.2 Process

The process studied is the cultivation of *Aspergillus oryzae* at the Fermentation Pilot Plant at Novozymes A/S. The Fermentation Pilot Plant is an experimental facility which among other functions serves to scale up new processes and test new strains and procedures. The filamentous fungus *Aspergillus oryzae* is widely used in biochemical industry due to its ability to produce extracellular enzymes. The

process studied in this work is production of the α-amylase sold as Fungamyl. The first step in the cultivation is preparation of an agar on which the strain is grown. This biomass is transferred to a shake flask and is then used to inoculate the seed cultivation tank. The seed cultivation tank contains an initial growth medium and only air is supplied. No pH control is carried out during this phase and the contents are transferred to the main cultivation tank when the pH decreases to a certain level. Further details regarding the seed cultivation tank are given in appendix A. Prior to the transfer the main cultivation tank is prepared and an initial amount of substrate is added. Initially the process is run in batch mode, only air and ammonia are supplied and off-gas is withdrawn[1]. The ammonia serves as both nitrogen source and to control the pH level. The purpose of the batch period is to generate a large amount of biomass in a short time, relatively little product is formed during this period. The high substrate concentration results in a high growth rate as well as a low production rate as the enzyme expression is substrate inhibited (chapter 2). When the initial substrate is exhausted, feed dosing is started and the process is changed to fed-batch mode. The depletion of substrate results in a rise in pH which serves as the trigger for the start of feed dosing. During the fed-batch phase the microorganism is substrate limited and the previously grown biomass produces the desired enzyme at a high production rate. Biomass is still produced in the fed-batch phase but due to substrate limitation this happens at a much lower rate than in the batch phase. A detailed description of the process operation in the main cultivation tank is given in appendix A.

3.2.1 Operational challenges

A number of factors must be considered during the cultivation process. The goal is to produce as large a quantity of enzyme as possible within the time period given. Due to the batch scheduling the duration of the batch is fixed on beforehand. As in every other chemical process one or more factors are limiting in the process. The real operational challenge is to have the most desirable limitation occuring at each point during the cultivation.

The substrate concentration is a very important factor for both biomass growth and enzyme production. High substrate concentrations favour biomass growth and repress enzyme production whereas low substrate concentrations favour enzyme production while having a very low biomass growth rate. The current strategy is therefore to produce a large amount of biomass in the batch phase having a high substrate concentration and shifting the process towards substrate limitation in the fed-batch phase ensuring high enzyme productivity of the produced biomass. The process is thus shifted from being limited by the maximal growth rate of the organism to being limited by the substrate supply rate. The substrate concentration should not become too low though because this leads to a reduced production rate of enzyme and under severe starvation parts of the biomass are likely to die which leads to greatly reduced performance. On the other hand a too high substrate concentration in the fed-batch phase will inhibit enzyme formation and lead to formation of excess biomass which decreases the product yield and increases the viscosity.

[1]In this work the term batch is used when no carbon substrate is fed to the process

Too high viscosity is another pitfall during the cultivation. High viscosity reduces oxygen transfer leading to limitation of oxygen transfer in parts of the biomass. This gives reduces productivity and even increases the high viscosity due to further formation of filaments (chapter 2).

Industrial production requires that the process is run in a very robust manner ensuring that the desired amount of product is produced in every batch. In practice this is achieved by running the process in a predefined manner according to a given recipe. Deviations from the recipe only happen if unusual or unexpected behaviour is observed by the process operators.

3.3 Process variables

The entire control and monitoring of the cultivations in the Fermentation Pilot Plant are carried out by the DeltaV control system. All the variables which are monitored and controlled are stored by the system in a PI (Plant Information) database. The data can later be extracted through the PIM process workbook as Excel files. Appendix B.3 provides more information on PI and PIM. The files are converted to Matlab format and are analysed further with this software. An overview of the variables recorded is given in table 3.1 and each of them is described in detail in appedix A.

3.3.1 Measurements available in pilot plant

The Fermentation Pilot Plant is meant to resemble the industrial scale production and the measurements available in the pilot plant are nearly the same as in the industrial plant. A description of each variable is given in the following. The units of each variable are given in table 3.1.

3.4 Data treatment

The pilot plant process data are collected either directly from the DeltaV system or as historical data from the PI database. For data collected directly from the DeltaV control system a sampling interval of 10 seconds has been used. More information on the DeltaV and PI system is given in appendix B.2. Data drawn from the PI database have been compressed and have a much lower sample rate and are subject to bias. Appendix B.3 gives a more detailed description of the data compression technique.

3.4.1 Truncation

The raw data contains periods which are not relavant for analysis, only the part of data with metabolic activity is included in the further analysis. Raw data often contains measurements made during tank preparation or after process termination. The starting point of the fermentation data is defined as the time point where the contents of the seed cultivation tank have been completely transferred to the main cultivation tank and the weight reading has been stabilised. In reality the transfer

Table 3.1. On-line and off-line measurements available from the Fermentation Pilot Plant

On-line variables	Unit
Cultivation time	hours
Dissolved O_2 tension	% of saturation
pH	-
Set point for feed flow rate	L/h
Measured feed flow rate	L/h
Integrated feed flow rate	L
Back pressure	barg
NH_3 flow rate	g NH_3/h
Integrated NH_3 flow rate	g
O_2 uptake rate (OUR)	moles/h
Integrated OUR	moles
CO_2 evolution rate (CER)	moles/h
Integrated CER evolution	moles
Broth weight	kg
Viscosity	cP
Refractive index	-
Respirative quotient (RQ)	-
Temperature	°C
Air flow	NL/min
Off-line variables	**Unit**
Enzyme activity	FAUF/g
Biomass, CTS	g/L
Ammonia concentration	mg/L
Laboratory pH	-

1g of enzyme corresponds to 9500FAUF, ie. 1FAUF corresponds to 0.105mg.

process only takes a few minutes and the data prior to the first stable weight reading are truncated. The very last data point is defined as the last point still having a stable weight reading. After this point the cultivation process is terminated by adding steam into the tank.

3.4.2 Outliers

Inspection of data reveals that a large number of outlying points is present. In this work an outlier is defined as a point which is clearly outside its normal range or even physically impossible. Outliers must be removed before further data processing such that erroneous information is not taken into consideration. It is important to be able to distinguish outliers from normal noise to avoid introduction of artefacts in the data.

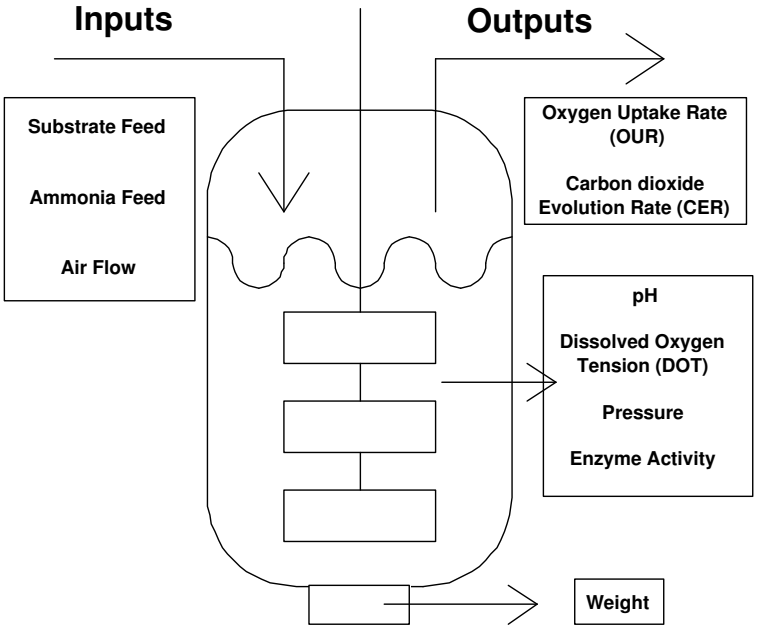


Figure 3.1. Sketch of the fermentor. Inputs to process are shown on the left hand side and outputs are shown on the right hand side

3.4.3 Missing data

In some cases blocks of missing data have been observed. One of the most common reasons is that the gas analysis measurements sometimes are out of order. The sample pipe can be blocked by cultivation broth or there can be other errors. In this case the calculated OUR and CER assume unrealistic values for a period of time. Fortunately the missing blocks are usally relatively short (5 to 8 minutes). In these cases linear interpolation has been made to fill out the missing block. In cases where a sensor has been showing erroneous readings throughout the entire cultivation the process variable concerned has been removed from the data set.

3.4.4 Bias

Bias in the measurements is a systematic error which poses problems for the data analysis and in worst case wrong conclusions can be drawn. Bias is introduced into the data sets in two ways:

Sensor bias The sensor can give biased readings due to wrong calibration or simply due to its age. This bias is difficult to detect but an indication of a biased sensor might be observed if the data is consequently outside the expected range. If redundant data is available data reconciliation might indicate biased sensors or other inconsistencies in data.

Database compression When a lossy compression technique [2] is employed bias might very well be introduced in the reconstructed data. Lossy compression

[2]Lossy compression is defined as a compression technique which is not able to transform the compressed data to be identical to the uncompressed data

is never able to the reconstruct the original uncompressed data and persistent errors are introduced in the data. The degree of bias introduced depends heavily on the compression technique employed (appendix B provides more information on the compression technique)

It is very important to take the possibility of bias into consideration even though it is generally difficult to detect and remove.

3.4.5 Filtering

Random errors and noise in the signal are removed by applying a fourth order Butterworth filter with cut-off frequencies fitted individually for each variable. The filtering has been done using the 'filtfilt' function in Matlab. This function filters the data first in the forward direction and then in the backward direction and thus eliminates the phase shift otherwise introduced. The need for filtering depends on the subsequent use of data. If the data is to be treated further in an algorithm which handles noise on its own one should be careful with prefiltering data because artefacts can easily be introduced.

3.5 Material balances

This section analyses material balances for the seven batches available from the Fermentation Pilot Plant. Material balances are a powerful tool to assess the quality of experimental data (Wang and Stephanopoulos (1983)). If the balances do not fit satisfactorily either one or more contributions have been neglected or the accuracy on the measurements is not sufficient. Combining more balances can in some cases provide additional information on deficiencies in the terms of the balance. An example of this is combining the carbon and degree of reduction balance to identify a compound missing in the analysis.

In this study balances have been set up for:

1. Carbon
2. Nitrogen
3. Degree of reduction

Unfortunately measurements for biomass are not available for three of the batches studied. For these batches carbon, degree of reduction and nitrogen have not been set up because the biomass contributes significantly in these balances. One batch, AFF1103, has an error in the weight reading which makes it impossible to set up proper balances. This batch is therefore left out of the material balance analysis.

The composition and degree of reduction of the constituents taken into account in the balances are given in table 3.2. For compounds containing carbon the composition and degree of reduction is reported per carbon mole.

Information regarding the composition of biomass and product as well as experimental procedures is found in appendix B. The degree of reduction of the compounds has been calculated according to the procedure given in Nielsen *et al.* (2003). The

Table 3.2. Composition and degree of reduction of the compounds included in the balances. Substrate accounts for both initial substrate and added feed dosing

Compound	Composition	Degree of reduction, κ
Biomass/C mole	$CH_{1.72}O_{0.55}N_{0.17}$*	4.11
Product/C mole	$CH_{1.50}O_{0.31}N_{0.25}S_{0.008}$	4.15
Substrate/C mole	$CH_{2x}O_x$	4
Mannitol/C mole	$CH_{7/3}O$	4.33
Malate/C mole	$CH_{3/2}O_{5/4}$	3
Oxygen/mole	O_2	-4
Carbon dioxide/mole	CO_2	0
Ammonia/mole	NH_3	0

* An ash content (c_{ash}) of 7.5% is assumed.

substrate composition is written as $CH_{2x}O_x$ in order to account for chains of glucose monomers. Each time a monomer is added to a chain one molecule of water is released. The ratio between H and O will therefore always be 2:1. This has an influence on the molar weight but not on the degree of reduction. Appendix A covers the calculation of the actual molar weight of the dosing.

3.5.1 Carbon balances

A carbon balance has been set up by considering the following terms:

1. Initially added substrate
2. Added feed dosing
3. Recovered biomass
4. Recovered enzyme
5. Recovered carbon dioxide

The concentration of substrate in the cultivation tank is not measured and is therefore not taken into account in the balance. In the batch phase there is a substantial substrate concentration whereas the concentration is usually negligible in the fed-batch phase due to carbon limitation. The amount of biomass transferred from the seed tank to the main fermenter is considered to be negligible and is also not taken into account in this analysis. The initial substrate and the added dosing are both derived from corn starch and it is assumed that the content of carbon which can be metabolised is the same in both cases. The calculation of the carbon concentration is given in appendix A.3.1. The density of the broth is assumed to be 1.03kg/L. The amount of substrate carbon which is transferred from the seed tank is disregarded as this amount ought to be very low as the micro organisms have consumed almost all the substrate in the seed stage. The compositions given in table 3.2 are used for the carbon balance. The total cultivation duration is divided

into four time periods and the balance below is given as the total number of carbon moles of each item at the end of each time period. The balance for time period, k, can subsequently easily be found as the difference between the balance at the end fo time period k and $k-1$. This approach has been used to calculate the yield coefficients in table 3.3.

$$C_{added}(k) = C_{in}(k) + C_{initial} \tag{3.1}$$

$$C_{in}(k) = feed_{acc}(k) \cdot \rho_{feed} \cdot \frac{c_{substrate}}{M_{substrate}} \tag{3.2}$$

$$C_{initial} = m_{substate,initial} \cdot \frac{c_{substrate}}{M_{substrate}} \tag{3.3}$$

$$C_{recovered}(k) = C_{biomass}(k) + C_{product}(k) + C_{out}(k) \tag{3.4}$$

$$C_{biomass}(k) = CTS(k) \cdot (1 - c_{ash}) \cdot \frac{m_{broth}(k)}{\rho_{broth} \cdot M_{biomass}} \tag{3.5}$$

$$C_{product}(k) = EnzA(k) \cdot \frac{m_{broth}(k)}{M_{product}} \tag{3.6}$$

$$C_{out}(k) = CER_{acc}(k) \tag{3.7}$$

$$\tag{3.8}$$

The calculated yield coefficients for conversion of substrate to biomass, product and carbon dioxide are given in table 3.3. The yield coefficients provide a convenient way to represent how the supplied carbon is distributed into products. The sum of yield coefficients is the observed recovery within each period.

Table 3.3. Yield coefficients for conversion of substrate to biomass, product and carbon dioxide respectively. The 3 batches with biomass measurements have been investigated

Batch	Period (h)	Y_{sx}	Y_{sp}*	Y_{sc}	$\Sigma(Y_{si})$
AFF1082	0-55	0.217	0.0476	confidential	0.967
	55-103	0.281	0.0592	confidential	1.156
	103-151	-0.106	0.0388	confidential	0.661
	151-197	0.698	0.0722	confidential	1.571
	0-197	**0.287**	**0.0548**	confidential	**1.103**
AFF1101	0-47	0.241	0.0394	confidential	0.940
	47-83	0.173	0.071	confidential	1.050
	83-119	0.136	0.0454	confidential	0.929
	119-167	0.438	0.0214	confidential	1.023
	0-167	**0.304**	**0.038**	confidential	**1.000**
AFF1102	0-47	0.239	0.0328	confidential	0.835
	47-95	0.064	0.0614	confidential	0.935
	95-143	0.207	0.0374	confidential	1.005
	143-180	0.102	0.0136	confidential	0.836
	0-180	**0.155**	**0.0388**	confidential	**0.907**

*Due to confidentiality the enzyme yields are scaled by a constant factor

The carbon balance for AFF1101 seems to fit perfectly at first sight. AFF1082 and AFF1102 close the total carbon balance with 110% and 90% respectively. The total recovery of AFF1082 is physically impossible and large variations in the observed recovery occur within the batch. Between 103 and 151 hours the total amount of biomass actually decreases which explains the negative yield coefficient. The yield coefficient for conversion of substrate to carbon dioxide on the other hand has relatively small variations between periods and batches indicating reliable measurements of this variable.

The yield coefficients related to procuction of enzyme vary somewhat but this can be expected due to the sensitivity of the production rate to e.g. concentrations of substrate and oxygen.

The large variations in the the biomass yields strongly suggest that this measurement is not reproducible and that the experimental procedure should be improved.

3.5.2 N balances

A nitrogen balance has been formulated considering the following nitrogen containing components:

1. Initial nitrogen from added salts
2. Added gaseous ammonia
3. Recovered biomass
4. Recovered enzyme
5. Measured ammonia concentration in the cultivation tank

The calculated yield coefficients for conversion of ammonia to biomass and product are given in table 3.4. The amount of nitrogen remaining in the cultivation tank compared to the amount supplied is given as Y_{nbroth}. The sum of these three terms is the total recovery within the period.

Table 3.4 shows that very large differences in recovered nitrogen are observed within the periods for all the batches. AFF1101 and AFF1102 do not have a complete set of off-line ammonia concentration measurements in the broth but for AFF1082 it seems that around 5% of the total supplied ammonia is found in the broth. Some of the large variations might be due to different amounts of ammonia accumulated in the broth but it is thought that uncertainty on the off-line measurements plays a big role.

3.5.3 Degree of reduction balances

A degree of reduction balance has been set up by considering the overall reacton given below:

$$biomass + products + CO_2 - substrate - NH_3 - O_2 = 0 \tag{3.9}$$

Based on this reaction the degree of reduction balance is written as:

$$balance = \kappa_{biomass} + \kappa_{products} - \kappa_{O2} - \kappa_{substrate} \tag{3.10}$$

Table 3.4. Yield coefficients related to nitrogen for the 3 batches containing biomass measurements

Batch	Period (h)	Y_{nx}	Y_{np}*	Y_{nbroth}	$\Sigma(Y_{ni})$
AFF1082	0-55	0.297	0.096	confidential	0.892
	55-103	0.575	0.178	confidential	1.426
	103-151	-0.222	0.119	confidential	0.361
	151-197	1.327	0.202	confidential	2.375
	0-197	**0.510**	**0.143**	confidential	**1.268**
AFF1101	0-47	0.233	0.0560	confidential	0.513
	47-83	0.311	0.188	confidential	1.249
	83-119	0.273	0.133	confidential	0.940
	119-167	1.057	0.0758	confidential	1.436
	0-167	**0.559**	**0.103**	confidential	**1.073**
AFF1102	0-47	0.284	0.0572	confidential	0.614
	47-95	0.157	0.223	confidential	1.185
	95-143	0.649	0.172	confidential	1.510
	143-180	0.426	0.0842	confidential	0.847
	0-180	**0.328**	**0.120**	confidential	**0.930**

*Due to confidentiality the enzyme yields are scaled by a constant factor

Carbon dioxide and ammonia are not included in the balance as their degree of reduction is zero.

The recovery of the degree of reduction equivalents is calculated as the recovered reduction equivalents divided by the ones added:

$$recovery = \frac{\kappa_{biomass} + \kappa_{products}}{\kappa_{O2} + \kappa_{substrate}} \tag{3.11}$$

The degree of reduction of the constituents is given in table 3.2.

The recoveries of the degree of reduction are given in table 3.5 where carbon and nitrogen recoveries are included for comparison.

Table 3.5 shows that the recoveries for the degree of reduction balance are typically even higher than those for carbon and nitrogen. However the recovery must be used with caution as the sum of the two terms in the denominator in equation 3.11 have opposite signs. The recovery therefore becomes very sensitive to even small errors in either oxygen or subtrate and this is why recoveries above 100% are observed. AFF1101 seems to close the degree of reduction balance as well as the carbon balance reasonably well.

3.6 Conclusion

This chapter has introduced the cultivation process analysed in this thesis and the experimental data obtained from the Fermentation Pilot Plant at Novozymes A/S. The material balances show that inconsistencies in the measurements are present. Both the carbon and nitrogen balance exceeds 100% recovery in some cases which

Table 3.5. Material balances for the 3 batches containing biomass measurements. The given numbers are recoveries in percent of the amounts added.

Batch	Period (h)	C-recovery (%)	N-recovery (%)	Red-recovery (%)
AFF1082	0-55	96.7	89.2	169.7
	55-103	115.6	142.6	140.3
	103-151	66.1	36.1	17.5
	151-197	157.1	237.5	226.1
	0-197	**110.3**	**126.8**	**172.8**
AFF1101	0-47	94.0	51.3*	105.1
	47-83	105.0	124.9*	110.9
	83-119	92.9	94.0*	107.4
	119-167	102.3	143.6*	98.8
	0-167	**100.0**	**107.3***	**105.0**
AFF1102	0-47	83.5	61.4	79.6
	47-95	93.5	118.5	83.5
	95-143	100.5	151.0	104.9
	143-180	83.6	84.7	152.4
	0-180	**90.7**	**93.0**	**96.8**

*AFF1101 does not contain measurements of the ammonia concentration in the broth.

is physically impossible. The observed yield coefficients for conversion of substrate strongly indicate that the largest uncertainties are to be found in the off-line biomass measurements. Furthermore the data from batch AFF1082 seems to be more corrupted than those of AFF1101 and AFF1102.

Many factors need to be considered when experimental data from an industrial system are analysed. The most important thing to keep in mind is the original purpose of the data. The analysed data is typically well suited for monitoring purposes and to detect irregularities in the process but when it comes to detailed analysis the quality might become insufficient.

Part II

Cultivations with standard measurements

4

Analysis of seed cultivation

4.1 Introduction

This chapter treats the seed cultivation of the Novozymes Fungamyl production process. The seed cultivation is carried out prior to the main cultivation to produce a certain amount of biomass which is transferred to the main cultivation tank. The seed cultivation is often overlooked in the entire enzyme production process but is nevertheless crucial as it determines the achievable performance of the entire main cultivation. If the biomass grown in the seed cultivation tank does not reach a sufficient concentration it will be impossible to reach a satisfactory enzyme yield in the subsequent main cultivation. The goal of the seed cultivation is to produce as much biomass as possible in the given time in order to ensure a sufficiently high biomass concentration in the main cultivation tank for being able to achieve the desired enzyme yield.

An estimation of the duration of the seed cultivation and the amount of biomass available from the seed cultivation would provide information on the expected progress of the main cultivation. For future operation this information could prove valuable as it makes it possible to have several different operating recipes for the main cultivation tank and switch between them depending on the progress of the seed cultivation.

There are several challenges in order to achieve reliable biomass estimations based on available online measurements. The length of the batch period is dictated by several factors including the state of the microorganism when it is transferred from the inoculation flask, the amount and composition of the growth medium, availability of oxygen and minerals as well as several other factors. The medium always has the same nominal composition but variations in eg. water and protein content will occur. The transfer criterium is defined as a certain pH drop to ensure that a consistent amount of biomass is transferred to the main cultivation tank. Available process variables like gas analysis and pH level can provide valuable information on the evolution of the process. This information can provide a basis to estimate the biomass concentration in the seed cultivation tank and thereby also the biomass amount transferred to the main cultivation tank. Therefore the purpose of this chapter is to explore the possibility of using the currently available measurements to estimate the biomass concentration throughout the seed cultivation.

For the entire analysis it is assumed that the amounts of minerals, nitrogen, etc. are adequate to support the growth of the fungus.

The chapter is structured as follows: Section 4.2 provides a description of the experimental data used in the analysis. Section 4.3 proposes a simple model for the biomass growth in the seed cultivation tank. Section 4.4 deals with applicability of

the pH probe to provide information on the growth of the microorganism. Section 4.5 investigates the possibility to predict the duration of the batch period in the main cultivation tank based on information available from the seed cultivation tank. Finally section 4.6 concludes the chapter.

4.2 Pilot plant seed data

This section gives a description of the seed tank data. This data could potentially provide important information on the initial conditions for the main fermentation. Eg. a biomass estimator could be used to determine how much biomass is actually transferred to the main fermenter. Furthermore it might become possible to determine how much biomass is actually active. Variables like carbon dioxide evolution, oxygen uptake rate and respiratory quotient could prove valuable for estimating the biomass concentration. Measurement of biomass concentration is not possible in the seed tank due to the medium which contains large substrate particles. The method for biomass measurement can not distinguish between biomass and other large particles which would make the measurement very unreliable.

The raw data has been truncated such that the start time is the point where the contents of the inoculation flask are transferred to the seed cultivation tank and the end time is the point where the contents of the seed cultivation tank are transferred to the main cultivation tank.

pH is not controlled in the seed cultivation unlike the main cultivation where ammonia is added to keep the pH at a certain level. In some cases the pH has been corrected by the operators during the seed cultivation. This has been done because the deviation between the off-line and on-line pH measurement has exceeded a threshold value. On the raw data this is seen as a sudden jump in the pH signal. In order to eliminate problems in the further data analysis the pH signal prior to the correction has been shifted by the same amount as the correction to eliminate the transition.

Due to technical difficulties with the online data collection all data for the seed tanks are taken from the PI database and are thus compressed data (see Appendix B for more information on the PI data collection system). The seed cultivation has been investigated for seven batches, namely: AFF1082, AFF1098, AFF1099, AFF1100, AFF1101, AFF1102 and AFF1103. All variables which are expected to have any influence on the cultivation have been collected and inspected, these variables are given in table 4.1.

Table 4.1. Variables available from the pilot plant seed cultivation which are investigated in this study

On-line variables	Unit
Dissolved O_2 tension	% of saturation
pH	-
Back pressure	bar gauge
O_2 uptake rate (OUR)	moles/h
Integrated O_2	moles
CO_2 evolution rate (CER)	moles/h
Integrated CO_2 evolution	moles
Weight	kg
Respirative quotient (RQ)	-
Temperature	°C
Air flow	NL/min

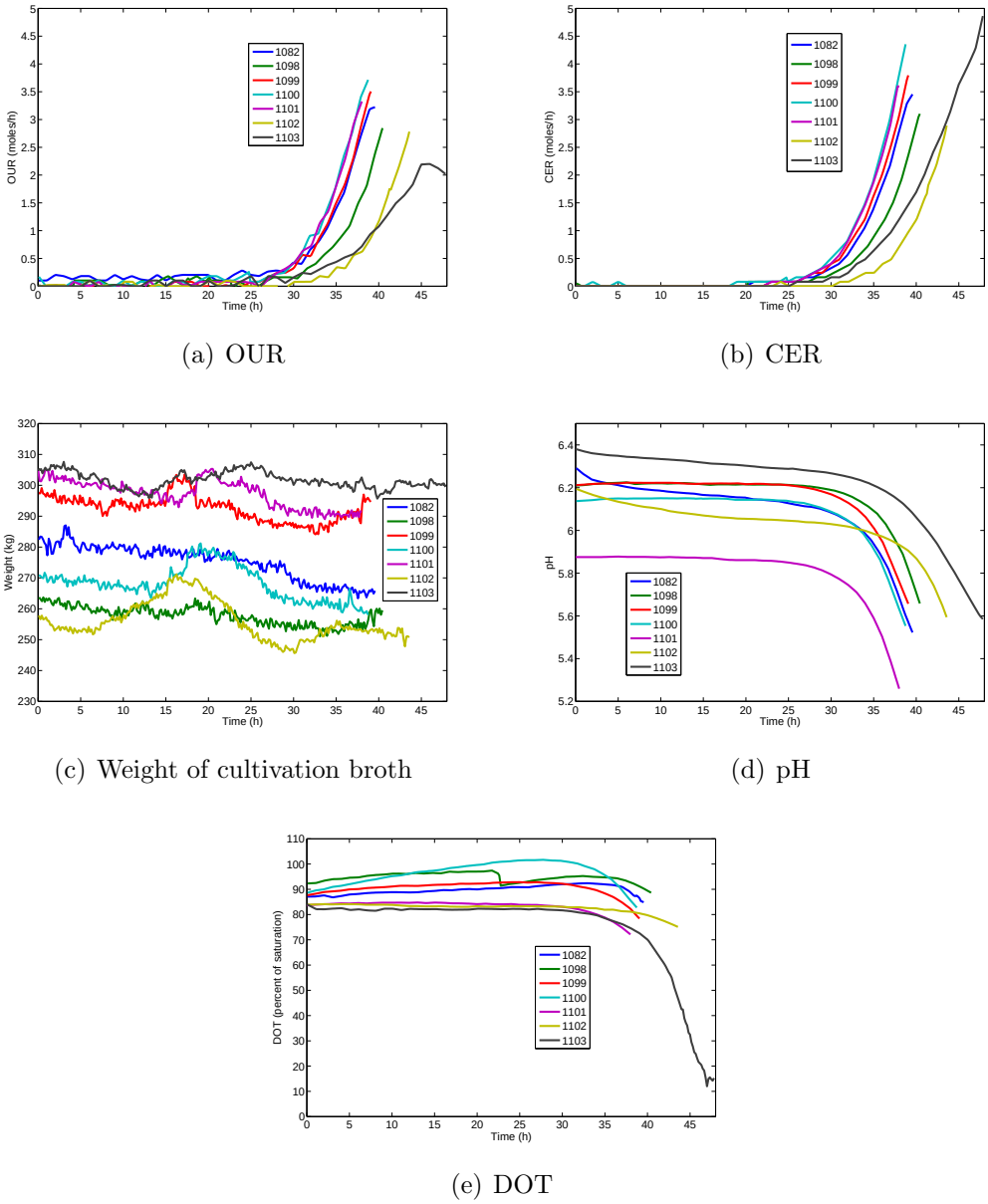
Five of the process variables have been shown in figure 4.2. The temperature and air flow rate are kept constant for all of the batches and have therefore not been shown. The gas analysis has relatively much noise and the calculated RQ value is rather noise sensitive as it is calculated as the ratio of CER and OUR and has not been shown here. The gas analysis system has a problem to detect very small values of OUR and CER. In the beginning of the cultivation the OUR and CER are very low and the measurements of the gas analysis fluctuate and are generally unreliable. This needs to be taken into account when parameter estimation is carried out. Stochastic differential equations are particularly well suited for parameter estimation in process data containing noise.

Figure 4.2 shows that the CER and OUR have a similar behaviour for the seven bathces investigated but in general CER has the highest value meaning that the RQ value is above one. The weight of the cultivation broth is generally decreasing due to water evaporation. However some batches have periods where the weight actually increases which is not readily explained.

The initial pH level is supposed to be the same for all batches but in practice it is not. In some of the batches this is due to the pH correction made during the cultivation. The initial DOT is around 90% and stays around this value during the major part part of the cultivation. Five to ten hours before the seed transfer DOT has a tendency to decrease a little bit which is due to an increased consumtion of oxygen. There is no indication that the process becomes oxygen limited at any point. AFF1103 is the only batch which has a very large decrease during the last part of the cultivation.

4.3 Growth model for the seed cultivation process

This section provides an attempt to develop a simple model for the seed cultivation based on the grey-box stochastic modelling framework described in chapter 2. To achieve this experimental data are combined with a simple first principle based



(a) OUR

(b) CER

(c) Weight of cultivation broth

(d) pH

(e) DOT

Figure 4.1. Overview of the seed cultivations. The five most important process variables are given.

model.

A model for the seed cultivation is proposed based upon a number of simplifying assumptions. The seed cultivation is run as a batch process with a high initial substrate concentration (Appendix A). This is known to lead to a high biomass growth rate and a low enzyme production rate (Chapter 2). It is assumed that the conditions for growth are unchanged during the seed cultivation process which leads to assumptions of constant specific growth rate, negligible product formation and constant yield coefficients. It is assumed the carbon contained in the substrate is converted to only biomass and carbon dioxide, which makes the carbon balance very simple. Furthermore it is assumed that the part of carbon utilisation related to

maintenance is negligible. Assuming that all the substrate is used up at the end of the seed cultivation phase the yield coefficients in the model can easily be estimated. The criterion for seed transfer is that the pH has dropped to below 5.3 from a value of around 6 in the beginning of the cultivation. The fact that pH is decreasing indicates that either acid is formed or base is utilised by the microorganism . This phenomenon has not been included in the present model.

The major assumptions used for modelling the seed cultivation are:

A1 No enzyme product or metabolites are produced

A2 Carbon in the substrate is converted into only biomass and carbon dioxide

A3 Yield coefficients can be assumed constant

A4 The growth rate is constant

A5 Maintenance is assumed negligible

The overall stoichiometry of the process is:

$$Y_{sx}X + Y_{sc}CO_2 + -S - Y_{so}O_2 = 0 \tag{4.1}$$

This stoichiometry combined with the assumptions mentioned leads to the following stochastic state space model.

State equations

$$dx = \mu x dt + \sigma_x dw_x \tag{4.2}$$

$$ds = -\frac{1}{Y_{sx}}\mu x dt + \sigma_s dw_s \tag{4.3}$$

Observation equations

$$OUR = Y_{xo}\mu x + e_{OUR} \tag{4.4}$$

$$CER = Y_{xc}\mu x + e_{CER} \tag{4.5}$$

4.3.0.1 Estimation of yield coefficients

The initial substrate provided in the seed cultivation tank consists of sucrose and soy beans in equal amounts by weight. The sucrose is well defined but the composition of the soy beans is complex and not known exactly. This makes it very difficult to estimate the content of carbon which can be utilised by the microorganism. The lack of biomass measurements further complicates the possibility to determine the yield coefficients for conversion of substrate into biomass and carbon dioxide. The method used in this work is to combine the estimated biomass at the very beginning of the main cultivation with the integrated CER for the seed cultivation. The amount of biomass initially present in the main cultivation tank is a known fraction of the amount present at the end of the seed cultivation process because the transferred amount is known. The final amount of biomass in the seed tank can therefore be

estimated if the estimated initial biomass amount in the main cultivation tank can be regarded reliable.

Under these assumptions an average yield coefficient between biomass and carbon dioxide can be estimated. The yield coefficients relating substrate to biomass and carbon dioxide are then estimated assuming that substrate is only converted to biomass and carbon dioxide.

$$Y_{sc} = Y_{sx}Y_{xc} \tag{4.6}$$

$$Y_{sx} + Y_{sc} = 1 \Leftrightarrow Y_{sx} + Y_{sx}Y_{xc} = 1 \Leftrightarrow Y_{sx} = \frac{1}{1 + Y_{xc}} \tag{4.7}$$

In chapter 5 it is found that not all the initial biomass estimates are reliable. Also it is found that the yield coefficients are not constant during the batch part of the main cultivation. In chapter 5 it is found that the most reliable estimate is the one for batch AFF1101. and this has been used to calculate the yield coefficients given in table 4.2. The yield coefficient between biomass and carbon dioxide, Y_{xc}, is found by dividing the total amount of carbon dioxide evolved in the batch by the estimated amount of biomass. The other two yield coefficients are calculated using equation 4.6 and 4.7.

Table 4.2. Estimated average yield coefficients for the AFF1101 seed fermentation. It is assumed that the substrate is converted to only biomass and carbon dioxide

Y_{xc}	Y_{sx}	Y_{sc}
0.610	0.621	0.379

4.3.1 Constant specific growth rate

Having estimated the yield coefficients related to substrate conversion the remaining parameters in the proposed model can now be estimated. The specific growth rate, initial amount of biomass and the yield coefficient between substrate consumption and oxygen production are estimated using the CTSM software (Kristensen and Madsen (2003)) with the model given by equations 4.1 to 4.5.

Data have been prepared for the CTSM software by subsampling the pretreated data to a sampling interval of 10 minutes (see Chapter 3 for information regarding data pretreatment). The original data set consists of approximately 2400 data points and with the sampling interval of 10 minutes this reduces to approximately 240 data points which reduces the computational burden significantly.

Table 4.3 shows estimation results of the initial biomass amount as well as the yield coefficient between substrate and oxygen uptake. The estimated amount of biomass is a little above 1 C-mole for all batches except for AFF1098 which is estimated to be eight times higher, the reason for this is most likely fluctuations in the gas analysis in the beginning of the process. The standard deviations are very reasonable compared to the parameter estimates. The yield coefficient between substrate and oxygen is between 0.339 and 0.375 for all the batches except for AFF1103 where Y_{so}

Table 4.3. Estimated initial amount of biomass and the yield coefficient between substrate consumption and oxygen consumption for the model given in equation 4.1 to 4.5. Both parameter estimates and uncertainties are calculated using the CTSM software

	$x_0\ (C-moles)$		$Y_{so}(mole/mole)$	
Batch	Estimate	Std. dev.	Estimate	Std. dev.
AFF1082	1.06E+00	1.26E-01	3.75E-01	3.71E-03
AFF1098	8.49E+00	8.28E-01	3.39E-01	2.19E-03
AFF1099	1.01E+00	7.62E-02	3.52E-01	1.07E-03
AFF1100	1.02E+00	1.45E-01	3.48E-01	1.56E-03
AFF1101	1.01E+00	7.01E-02	3.62E-01	1.51E-03
AFF1102	1.01E+00	6.41E-02	3.70E-01	1.62E-03
AFF1103	1.02E+00	2.41E-01	2.18E-01	1.62E-03

is 0.218. Again the standard deviations are small compared to the estimates. These values are all lower than Y_{sc} thus the average respirative quotient is above one. The estimate of Y_{so} for AFF1103 is very far from the remaining six batches which could indicate a problem with the data for this batch.

Table 4.4. Estimated specific growth rate and standard deviation for the state noise term related to the biomass growth equation (4.2). Both parameter estimates and uncertainties are given

	$\mu\ (1/h)$		$\sigma_x\ (C-moles)$	
Batch	Parameter estimate	Std. dev.	Parameter estimate	Std. dev.
AFF1082	8.48E-03	1.81E-04	7.46E-01	4.77E-02
AFF1098	9.58E-03	1.23E-04	4.00E-01	2.25E-02
AFF1099	8.98E-03	1.10E-04	4.88E-01	2.60E-02
AFF1100	9.14E-03	1.36E-04	6.91E-01	3.76E-02
AFF1101	9.94E-03	1.27E-04	4.42E-01	2.48E-02
AFF1102	9.79E-03	1.61E-04	5.00E-01	2.57E-02
AFF1103	5.54E-03	1.08E-04	1.11E+00	5.50E-02

Table 4.4 lists estimated parameters and uncertainties for the estimated specific growth rate and the standard deviation of the noise related to the state equation describing biomass growth (equation 4.2).

The estimates of specific growth rate are all very close to each other (around $9 \cdot 10^{-3}$) except for AFF1103 for which the growth rate estimate is only $5.54 \cdot 10^{-3}$. The standard devitions are all relatively low compared to the estimates. However a closer look at the estimation results shows that the estimates for the state noise term of the biomass formation equation (σ_x in equation 4.2) is significant for all the batches (table 4.4). This suggests that equation 4.2 is not able to describe the biomass growth correctly. A possible explanation for the results is that the growth rate is not constant but that the estimation converges to a solution with a constant

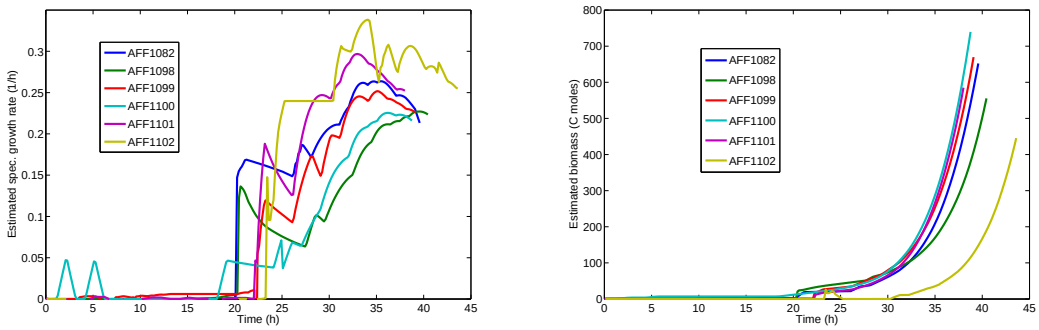
specific growth rate and a large process noise term. As a consequence it is decided to investigate the specific growth rate in more detail.

4.3.2 Varying specific growth rate

The results of the previous subsection indicated that the specific growth rate can not necessarily be assumed constant. A varying specific growth rate is accounted for by adding it as an additional state in the model given previously. The evolution of the specific growth rate is estimated by the state estimator (Extended Kalman Filter) included in the CTSM package. The model is identical to the one given in equations 4.1 to 4.5 except for the added equation equation (4.8),

$$d\mu = 0 + \sigma_\mu dw_\mu \tag{4.8}$$

which simply states that the specific growth rate is modelles as constant hence all variation is captured in the process noise term. The state estimates have been generated using the 'Generate prediction data' function in CTSM. This function uses the Extended Kalman filter to calculate $\hat{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k}}$, the estimated states are shown in figure 4.2.



(a) $\mu(t)$ (b) $x(t)$

Figure 4.2. Estimated specific growth rate as function of time (a) and estimated amount of biomass (b). The estimation is based on the model given in equations 4.1 to 4.5 and 4.8 combined with measured gas analysis data subsampled to a 10 min sample interval

Figure 4.2a shows the estimated specific growth rate for the 6 batches AFF1082 throgh AFF1102. AFF1103 has not been included in the plots as the results are believed not to be trustworthy.

The results show that there is a long initial lag period of around 17-23 hours where the growth rate is close to zero. The specific growth rate then increases rapidly and reaches a maximal value of between 0.22 and 0.34/h which corresponds well to values found in literature (eg. Spohr *et al.* (1997)). Depending on the batch the maximum specific growth rate appears between 33 to 40 hours. The decrease in specific growth rate in the very last part of the cultivation might very well be due to substrate limitation. This will however not be investigated further due to the lack of precise information on the substrate composition.

Figure 4.2b shows the total estimated amount of biomass given in C-moles. In the lag period the biomass is virtually zero and rapid growth is observed in the last half of the cultivation. The final amounts of biomass vary somewhat between bathces.

It is interesting to note that the lag period and total cultivation time are relatively similar between batches (Figure 4.2a). This clearly shows that cultivations to a large extent can be run in a reproducible manner.

The estimates of the specific growth rate are very low during the first 20h, i.e. too low to provide discernible informaton in the gas measurements. Thereafter the specific growth rate is estimated to attain a value between 0.20 to 0.30 h^{-1} over the next 15 to 20 hours until pH drops and the seed cultivation is terminated.

The estimates of the specific growth rate are clearly affected by the noise present on the CER and OUR data which limits the possibility to investigate any functional relationships in data.

4.4 Estimation of biomass concentration based on pH measurement

This section investigates to which extent the pH signal can provide information on the specific growth rate and amount of biomass in the seed cultivation. A simple biomass estimator can in principle be developed based on the gas analysis but as shown in section 4.3 this is associated with challenges regarding the reliability of the measurements. Except for calibration errors the pH signal is generally regarded as relatively reliable and using it as a biomass estimator therefore contributes interesting perspectives.

The fact that pH is decreasing towards the end of the seed cultivation indicates that acid (protons) is accumulating. *Aspergillus oryzae* takes up ammonia as the ammonium ion (Griffin (1994)) and excretes a proton in this process. In the seed cultivation no ammonia is present so the nitrogen demand is covered by the nitrogen present in the soya added. As the protons are not formed as a result of ammonia uptake they could be a result of formation of e.g. organic acids. In case the proton formation is proportional to the formation of biomass the pH can be used as an indicator of biomass growth. In order to investigate whether this relationship is present in the following the assumption is made that the ratio between proton excretion and biomass formation is constant. This leads to the following stoichiometric model:

$$Y_{sx}X + Y_{sc}CO_2 + Y_{sH}H^+ - S - Y_{so}O_2 = 0 \tag{4.9}$$

Based upon the measured pH the concentration of protons has been calculated as $c_{H^+} = 10^{-pH}$, the total number of protons formed during each of the seed cultivations is shown in figure 4.4.

Figure 4.4 shows that there is an unexpected deviation in the calculated number of protons produced in the different batches. The biomass estimates based on gas analysis (figure 4.2b) were very similar for all batches except AFF1102. The explanation for the deviation in this case is most likely that the pH signal is not completely consistent (figure 4.2d) combined with noise on the mass measurement (figure 4.2c). The pH signal does not follow the same trajectory during the first

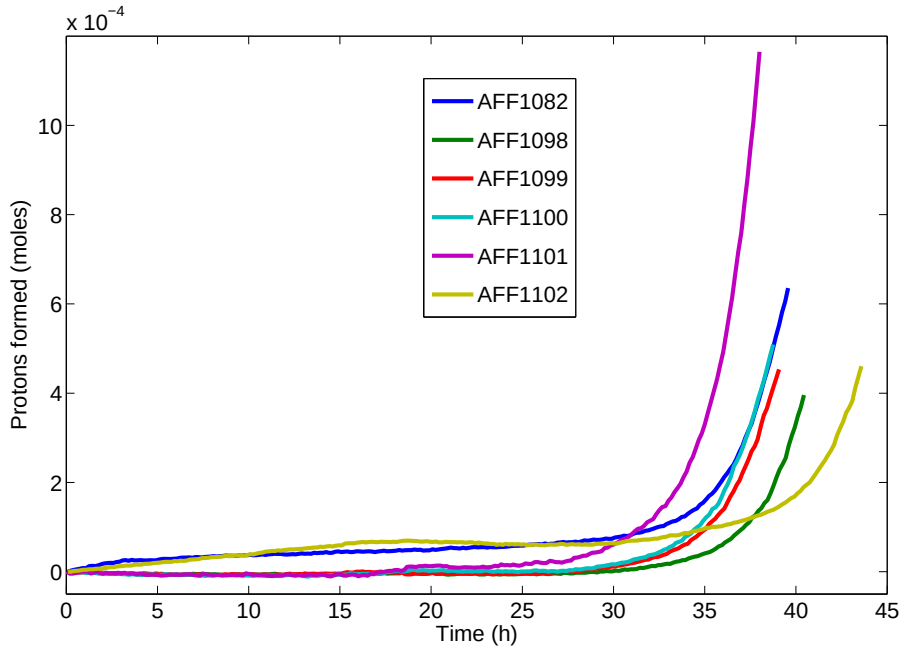


Figure 4.3. Estimated number of protons formed (moles) in the different seed cultivation batches

half of the cultivation which leads to large differences in the calculated number of protons. On figure 4.2d it also seems that the pH signal has different biases. Since the pH units are logarithmic this also has a large influence on the proton estimation. These observations indicate that the current setup and calibration of the pH sensor is not sufficiently accurate to make reliable estimations. Keeping this in mind average values of Y_{xH} are calculated based on the total estimated biomass formed in each batch combined with the total proton consumption for each of the batches. The results are given in table 4.4.

Table 4.5. Y_{xH} for seed fermentation

Batch	$Y_{xH}/10^{-7}$ (C mole/mole)	$Y_{sH}/10^{-7}$ (mole/mole)
AFF1082	9.75	6.05
AFF1098	7.13	4.43
AFF1099	6.77	4.20
AFF1100	6.38	3.96
AFF1101	19.9	12.4
AFF1102	10.3	6.40

Table 4.4 gives both Y_{xH} and Y_{sH} for the seven batches investigated. The moles of protons excreted per C mole of biomass formed is in the order of one to one million. The average yield coefficient, Y_{xH}, is calculated to vary between $6.38 \cdot 10^{-7}$ and $19.9 \cdot 10^{-7}$ between batches.

The observations indicate that a constant yield coefficient between proton excre-

tion and biomass formation can be excluded with the present measurements.

4.5 Estimation of batch period in main cultivation tank based on seed tank data

This section investigates if the estimated final biomass concentration in the seed cultivation tank can be used to predict the duration of the batch phase in the main cultivation tank. This can yield interesting perspectives with respect to optimising the productivity of the main fermentation because it can provide a possibility to use different feeding strategies depending on the progress and duration of the batch phase.

Table 4.6. Duration of the batch period of the main cultivation compared to the initial starch amount in the main cultivation tank as well as the estimated amount of biomass transferred to the main cultivation tank

Batch	Initial starch (kg)	Biomass transferred (C moles)	Batch period (h)
AFF1082	45	306.7	24.7
AFF1098	50	256.9	29.7
AFF1099	50	272.8	27.1
AFF1100	50	371.4	27.1
AFF1101	4	240.9	9.37
AFF1102	4	213.2	10.1

Table 4.6 lists the initial amount of starch in the main cultivation tank, the estimated amount of biomass transferred from the seed cultivation tank to the main fermentation tank as well as the duration of the batch period in the main cultivation tank for the six batches investigated. The estimation of the amount transferred to the main tank is based on the model given by equations 4.1 to 4.5 and 4.8 combined with measured gas analysis data subsampled to a 10 min sample interval as explained in section 4.3.2. The estimated final amount of biomass in the seed tank has been combined with the ratio between the mass transferred to the main cultivation and the mass of the seed cultivation broth. The numbers in table 4.6 are therefore less than what is indicated by figure 4.2. The horisontal line between AFF1100 and AFF1101 is to indicate that different initial amounts of starch have been used in the investigated cultivations, 45/50 kg and 4 kg. The initial amount of starch in the main cultivation tank has a large influnce on the duration of the batch period. For AFF1101 and AFF1102 it is seen that AFF1101 has the higher transferred biomass amount and hence the shorter batch period. For the batches run with the higher initial substrate amount the results are not as clear. AFF1099 and AFF1100 have very different transferred biomass amounts but identical batch durations. AFF1098 has the longest batch duration and the lowest transferred biomass which agrees well with the expectation.

It must be noted that the biomass amounts are all estimated under a number

of assumptions and no actual measurements of the biomass are available. Having a better estimate of the biomass it can therefore not be excluded that a better prediction of the batch duration can be made.

Estimates of the initial amount of biomass in the main cultivation tank are given in chapter 5 and estimated values are found in table 5.3. These values should correspond well to the estimates of the amount of biomass transferred to the main cultivation tank based on the gas analysis measurements in the seed cultivation (the values in table 4.6). However, the values estimated in this chapter are in the range 10-20 times large than those given in table 5.3. The latter estimates are considered more reliable as some experimentally measured biomass amounts have been available during the batch cultivation. This finding indicates that the yield coefficient, Y_{sx}, can not be assumed to have the used value throughout the seed cultivation. A possible explanation is a significantly different metabolism due to the different growth medium which leads to formation of large amounts of biproducts.

4.6 Conclusion

The analysis of the seed cultivation reveals some interesting observations. Generally a lot of noise is present in the experimental data which complicates modelling and the ability to make accurate estimations. The most important observations made in this chapter are summarised in the following.

The specific growth rate has been estimated by applying the grey-box stochastic modelling framework. A lag period of 17 to 23h is observed and hereafter the specific growth rate increases rapidly to reach a value of 0.22 to 0.34/h depending on the batch. Towards the end of the period the specific growth rate decreases slightly.

The pH signal can provide an indication of the biomass growth but can not be used as a reliable biomass estimator based on the available experimental data.

The duration of the batch period in the main cultivation tank is correlated to the amount of biomass transferred from the seed cultivation tank but due to uncertainties in the estimations an accurate prediction can be made at the current point.

5

Analysis of main cultivation

5.1 Introduction

This chapter concerns the main cultivation of the Novozymes Fungamyl production process. The main cultivation is started by transferring the contents of the seed cultivation to the main cultivation tank which has previously been filled with a certain amount of initial starch. The beginning of the cultivation is run as a batch process and at the point where pH rises above a certain value the dosing feed is initiated. The goal of the main cultivation is to produce as much enzyme as possible within the given time. Unlike the seed cultivation it is not a goal in itself to achieve a high biomass concentration even though this is related to achieving a high enzyme yield.

Being able to analyse and synthesise the available online measurements provides a means to gain a deeper understanding of the progress of the cultivation and being able to control the process by manipulating inputs.

The analysis of the progress of the main cultivation is subject to a number of limitations. In chapter 3 it was found that the material balances are not as consistent as one could wish for. This means that the experimental data available are not accurate and a certain degree of uncertainty must be expected on the estimated results. Also variations occur in factors like medium composition, morphology of the microorganism and other conditions determining biomass growth and enzyme production. Despite the deficiencies in data a preliminary analysis of the data is performed in order to get new insight and help to plan further experimental investigation.

In this chapter only the analysis of the batch phase is given. The analysis of the fed-batch phase was not feasible due to lack of reliable measurements of the biomass concentration. An overall carbon balance has been applied to estimate the amount of biomass but successful results could unfortunately not be obtained. In this carbon balance it has been assumed that the supplied substrate balances the substrate uptake of the micro-organism which is not necessarily true. The analysis of the batch part is feasible because the enzyme production rate is assumed negligible in this phase of the process. This allows for estimation of biomass by use of the carbon balance.

In this section experimental data from the pilot plant at Novozymes A/S (introduced in chapter 3) will be analysed using the stochastic grey-box modelling framework described in chapter 2. This is done by combining experimental data with a simple first principle based model. The approach is similar to the one used in chapter 4.

The chapter is structured as follows: Section 5.2 proposes a simple model for the biomass growth in the main cultivation tank. Section 5.3 extends this model to

account for a varying specific growth rate. Section 4.6 concludes this chapter.

5.2 Analysis of batch phase assuming a constant specific growth rate

This section analyses the batch phase of the main cultivation process. The reason for only considering the batch phase is that a simple carbon balance can be formulated to provide information on the biomass growth. The initial amount of substrate is known and at the end of the batch phase this amount is assumed to have been completely depleted. The enzyme production rate is known to be relatively low during the batch phase due to the high substrate concentration and as rough estimate it is set to zero. Some enzyme is formed but the first measurement of enzyme activity is usually made after the end of the batch phase and information on the enzyme activity during the batch phase is therefore not available. These assumptions make it possible to set up a simple carbon balance and estimate the yield coefficients related to conversion of the carbon source. The biomass is assumed to be uniform with respect to composition and morphology and it is all assumed to be active and to express the same metabolic behaviour.

Depletion of the initially supplied substrate is seen as an increase in pH. Initially pH has a value of 6 which is controlled by addition of ammonia. When pH rises above 6.3 the dosing feed is automatically initiated and this defines the end of the batch phase. In this investigation the end of the batch phase is defined as the point where pH starts to increase.

AFF1103 is excluded in the further analysis because the carbon balance is not realistic for the batch phase because the number of moles of carbon dioxide exceeds the number of moles of initial substrate (see chapter 3). This is either due to poor CER measurements or that the initial amount of starch was not the one given in the process journal.

The major assumptions used for modelling the batch phase are given below:

A1 No enzyme product or metabolites are produced

A2 The substrate is converted into only biomass and carbon dioxide

A3 Yield coefficients can be assumed constant

A4 The growth rate is constant

A5 Maintenance is assumed negligible

A6 The amount of substrate transferred from the seed tank is negligible

With these assumptions the overall stoichiometry of the reaction becomes:

$$Y_{sx}X + Y_{sc}CO_2 - S - Y_{sn}NH_3 - Y_{so}O_2 = 0 \tag{5.1}$$

The assumptions lead to the kinetics given in the following stochastic state space model.

State equations

$$dx = \mu x dt + \sigma_x dw_x \tag{5.2}$$

$$ds = -\frac{1}{Y_{sx}} \mu x dt + \sigma_s dw_s \tag{5.3}$$

Observation equations

$$OUR = Y_{xo} \mu x + e_{OUR} \tag{5.4}$$

$$CER = Y_{xc} \mu x + e_{CER} \tag{5.5}$$

5.2.1 Estimation of yield coefficients

Assuming that all substrate is converted to biomass and carbon dioxide the carbon balance can be set up as:

$$Y_{sc} + Y_{sx} = 1 \Leftrightarrow Y_{sx} = 1 - Y_{sc} \tag{5.6}$$

The yield coefficient for conversion of substrate into carbon dioxide is estimated by assuming that all the initial substrate has been used completely at the end of the batch phase. The estimated yield coefficients are given in table 5.1.

Table 5.1. Estimated yield coefficients for the batch phase assuming no enzyme is formed. The initial amount of starch in the cultivation tank is also given

Batch	Initial starch (kg)	Y_{sc}	Y_{so}	Y_{sx}
AFF1082	45	0.448	0.425	0.552
AFF1098	50	0.417	0.391	0.583
AFF1099	50	0.424	0.397	0.579
AFF1100	50	0.444	0.416	0.556
AFF1101	4	0.291	0.260	0.709
AFF1102	4	0.195	0.178	0.805

The estimated values for Y_{sc} and Y_{so} are considered as being reliable because only initial starch, integrated CER and integrated OUR are used in the calculation. The calculated values for Y_{sx} include biomass as well as all other products, therefore the real values for Y_{sx} will definitely be lower.

The calculations show that batches having a low initial substrate concentration seem to have a higher yield coefficient for biomass, i.e. relatively less carbon is converted to carbon dioxide. This is unexpected as the metabolism is assumed to be independent of the concentration of starch. Two apparent explanations for this observation are:

1. The conversion of substrate is more efficient at low biomass concentrations which could be connected to limitations in part of the biomass

2. A significant amount of substrate is transferred from the seed tank. This would give a high apparent biomass yield for batches with low initial substrate level.

An explanation can not be given by the current analysis but will be further investigated in the following.

5.2.2 Estimation of constant specific growth rate

The specific growth rate and initial amounts of biomass have been estimated using the model given by equation 5.2 to 5.5 using CTSM.

Table 5.2. Estimated specific growth rate and initial amount of biomass for the model given in equation 5.1 to 5.5. Both parameter estimates and uncertainties are given. The horisontal line is to indicate that different amounts of initial starch have been used (see table 5.1)

	μ $(1/h)$		x_0(C moles)	
Batch	Estimate	Std. dev.	Estimate	Std. dev.
AFF1082	2.66E-02	1.40E-02	156	115
AFF1098	2.54E-02	1.26E-02	122	103
AFF1099	2.70E-02	1.45E-02	143	134
AFF1100	2.91E-02	1.45E-02	172	124
AFF1101	2.438E-01	6.89E-03	9.72	0.73
AFF1102	2.337E-01	8.46E-03	12.2	1.26

Table 5.2 shows the estimated parameters and uncertainties. The estimates of the specific growth rate are in the range between $2.54 \cdot 10^{-2}$ and $2.33 \cdot 10^{-1}$. The higher growth rates are observed for the batches with a low amount of initial substrate. The most likely explanation seems to be that this is related to limitations present at higher biomass concentrations. The standard deviations are very high for the first four batches (the ones with high initial substrate concentration) which strongly suggests that the specific growth rate can not be assumed constant. For the latter two cultivations the standard deviations are very small compared to the estimates and it is not unreasonable to assume constant growth rates here.

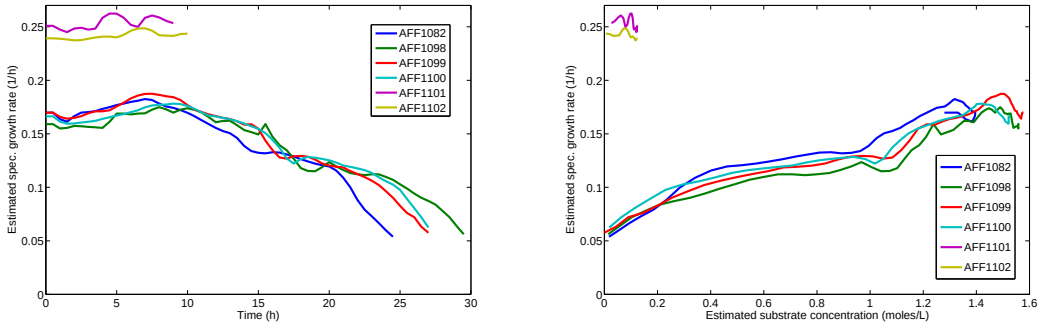
The estimates of the initial amount of biomass vary by a factor of almost 18 (9.72 to 172) and the standard deviations are also relatively large. The variations in the estimated initial amounts of biomass are large and it seems that the first four and last two are close to each other. This is unexpected as the amount of initial biomass is completely independent of the initial amount of starch. The explanation must be found in the assumption of a constant growth rate which introduces an artificially high value for the initial biomass amount.

5.3 Analysis of batch phase introducing a varying specific growth rate

A varying specific growth rate is investigated by adding it as an additional state in the model and subsequently performing state estimation. The model is identical to the one given in equation 5.2 to 5.5 to which the equation below has been added:

$$d\mu = 0 + \sigma_\mu dw_\mu \tag{5.7}$$

The state estimates have been generated based upon the measured data and the estimated models using the 'Generate prediction data' function in CTSM. This function uses the Extended Kalman Filter to calculate $\hat{\boldsymbol{x}}_{\boldsymbol{k}|\boldsymbol{k}}$, the estimated states are given in figure 5.1.



(a) $\mu(t)$ (b) $\mu(s)$

Figure 5.1. Estimated specific growth rates as function of time (a) and estimated substrate concentrations as function of time (b). The estimation is based on the model given in equations 5.1 to 5.5 and 5.7

Figure 5.1(a) and (b) show the estimated specific growth rate for the 6 batches which have been investigated. The estimated specific growth rates and substrate concentrations respectively have been plotted as functions of time. As expected it is evident from these figures that the specific growth rate can in general not be assumed constant.

The initial biomass amounts have been reestimated and are given in table 5.3.

As can be seen in table 5.3 the estimated initial biomass amounts differ significantly from the ones found in the first iteration (table 5.2). This is especially the case for the four first batches with the higher initial substrate concentration. Also the standard deviation for the estimate has been reduced significantly.

Two interesting observations are made from figure 5.1(a) and (b). The first one is that the specific growth rate starts at a much higher value in the batches with low initial substrate concentration (AFF1101 and AFF1102) than in the batches having the standard initial substrate concentration. The specific growth rate starts close to 0.25/h for the former and only around 0.16/h for the latter (figure 5.1(a)). These phenomena can not be explained by a simple growth model with Monod like dependencies of substrate concentration on the specific growth rate because

Table 5.3. Estimated initial amounts of biomass for the second model iteration. The model given in equation 5.1 to 5.5 and 5.7 has been applied. Both parameter estimates and uncertainties are given. The horisontal line is to indicate that different amounts of initial starch have been used (see table 5.1)

	x_0 $(Cmoles)$	
Batch	Estimate	Std. dev.
AFF1082	29.2	4.48
AFF1098	19.4	4.11
AFF1099	23.3	7.67
AFF1100	26.8	5.04
AFF1101	9.39	0.73
AFF1102	11.9	1.21

the observation made is that higher substrate concentrations lead to lower specific growth rates. The second observation is made in figure 5.1(b). It is seen that the specific growth rate is up to five times higher for AFF1101 and AFF1102 than for the other four batches at low substrate concentrations towards the end of the batch phase (0.25/h compared to 0.05/h at the left hand side of figure 5.1(b)). At this point in the cultivation the primary difference between the two groups of cultivations is different biomass concentrations. The biomass concentration in itself is not expected to have any direct influence on the specific growth rate but rather an indirect influence through some other property affected by the biomass concentration.

These two observed behaviours of the specific growth rate can possibly be explained as the substrate and biomass concentrations having an indirect influence on the specific growth rate through some other property. One relevant property is viscosity. Viscosity has a large effect on mass transfer rates which can severely affect the specific growth rate if in fact mass transfer is a limiting factor.

5.3.1 Dependence of viscosity on the specific growth rate

Figure 5.3.1 shows the viscosity for the batches investigated for times up to 25h. Batches AFF1099 and AFF1100 are not included in the figure as viscosity was not measured in these batches.

Figure 5.3.1 clearly shows that the batches with low initial substrate concentration (AFF1101 and AFF1102) have a considerably lower viscosity than the ones with standard concentrations. This suggests that the concentration of starch has a very large impact on the viscosity of the cultivation medium. The dependence of medium viscosity on the specific growth rate suggests that the prevalent limitation is mass transfer.

It must be kept in mind that the estimations of the specific growth rates are based on the assumption of constant yield coefficients. If the metabolism and thereby yield coefficients change during the batch phase this will affect the apparent specific growth rate. This has the effect that the calculated yield coefficients should be regarded as average yield coefficients for the observed period. The real yield

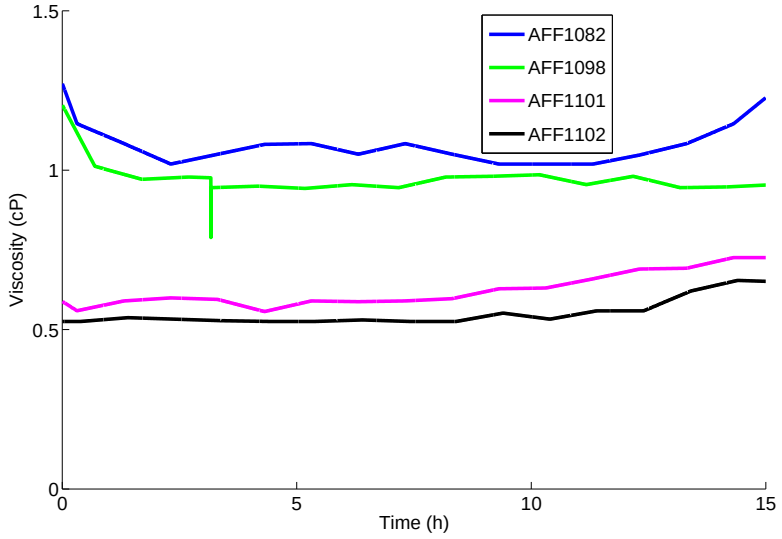


Figure 5.2. Viscosity measurement for four of the investigated batches plotted as function of time

coefficients at the beginning of the cultivation are expected to be independent of the starch concentration which means that the differences in average yield coefficients result from changes in metabolism occuring later in the batch phase. This means that some of the difference in the initial specific growth rate might very well be explained by changes in metabolism later in the batch phase.

One such change in metabolism is that enzyme starts to form. Enzyme formation has been disregarded in this analysis due to lack of measurements but if the production rate changes significantly at some point of the cultivation the yield coefficients will be affected. Another change in metabolism is that a limitation starts to dominate. Initially the process is limited by the maximal specific growth rate of the microorganism and no mass transfer limitations are present. The change in metabolism suggests a limitation in either available substrate or in mass transfer. The substrate concentration is expected to be sufficient to support growth during the batch period (except the very end) hence mass transfer limitation seems to be the most likely explanation. A possible mechanism leading to limitation is formation of large pellets. In other cultivation processes it is often observed that pellets become so large that limitation in oxygen and/or substrate transfer occur (Nielsen and Villadsen (1992), Moreira *et al.* (1996)). The type of limitation depends on the bulk concentration of oxygen and substrate respectively as well as pellet size, porosity, rate of oxygen/substrate uptake etc. (Carlsen (1994)).

The calculated specific growth rates for the batches with standard substrate concentration (figure 5.1(a)) increase at the beginning of the cultivation and start a decrease around 8-10 hours. Cultivations experiencing pellet growth often show two such operating regimes (Poulsen (2005)), one where the specific growth rate is more or less constant and one where it decreases rapidly. This phenomenon is not observed in this analysis but could very well be masked by varying yield coefficients.

The present data is not sufficient to answer the questions regarding formation of

pellets, changes in metabolism etc. An experiment with additional measurements is necessary to uncover the responsible mechanisms.

5.4 Conclusion

In this section an analysis of the main cultivation has been performed. Only the batch phase has been investigated due to lack of reliable biomass measurements in the fed-batch phase, furthermore the carbon balance does not fit very well for this phase. Still a number of important observations have been made in this chapter. One of the major observations is that at a certain biomass concentration the specific growth rate starts to decrease. This can be explained by a limitation occurring at this point, possibly due to formation of pellets.

The concentration of initial starch in the cultivation tank has a large influence on the viscosity of the medium. This seems to lead to a mass transfer limitation for the batches run with standard initial starch concentrations. The batches having a low initial starch concentration did not experience this limitation and the substrate is also utilised more efficiently in these batches. The yield coefficient for conversion of substrate to biomass is higher for the batches with a low initial substrate concentration suggesting that a less efficient metabolism sets in at a certain point for the batches having the standard initial substrate concentration.

The observations made in this chapter have led to the decision to carry out an additional experiment based on the standard recipe but having a high frequency of off-line sampling and a more thorough subsequent analysis of the samples.

Part III

Cultivation with additional measurements

6

Experimental investigation for characterisation of detailed cultivation

6.1 Introduction

The initial analysis of a number of pilot plant cultivation batches from Novozymes A/S showed that proper first principles engineering modelling was not possible due to a lack of understanding of the metabolism and large uncertainties on especially off-line laboratory measurements. The growth during the batch phase expressed a behaviour which is not readily explained by standard growth models. In order to obtain the information necessary for further modelling of the fermentation process an experiment has been carried out with a detailed measurement program to investigate different aspects of the cultivation in greater detail than in previous experiments.

As discussed in section 2.5.1 the morphology is a combination of pellets and freely dispersed hyphae. The latter can further be divided into free hyphae and clumps which are agglomerates of freely dispersed hyphae but with much lower density than actual pellets. This complicates the identification of the active metabolism as limitations inside the pellets are likely to exist. In practice several different metabolisms may be active at the same time (eg. no limitation, substrate limitation and oxygen limitation) and the overall observation is a combination of those. It is very important to take these aspects into consideration when process variables are investigated and relationships such as yield coefficients are determined. Variations in overall stoichiometry during the cultivation exhibit a combination of changes in metabolism and distribution between different morphologies. Segregated models (eg. Agger *et al.* (1998)) are able to account for regions having different properties but require rather detailed information of the morphology.

Hence the purpose of the experiment is to characterise the cultivation with respect to metabolite formation and consumption as well as attempting to gain a deeper understanding of the mechanisms determining the progress of the cultivation.

This chapter is structured such that a description of the experimental design and sample withdrawal policy used is provided in section 6.2. In section 6.3 it is found that the entire duration of the cultivation can be divided into operating regions with distinct features. Section 6.4 analyses the material balances to assess the quality of the obtained experimental data. Section 6.5 provides an analysis of the changes in morphology occuring in the process. A conclusion is given in section 6.6.

6.2 Experimental design

The cultivation has been run using the standard recipe and a standard feeding profile, the only difference is the frequency of sample withdrawal and the subsequent analysis work. At 150 hours a correction has been made to the feed dosing profile due to the occurence of severe oxygen limitation. At 200 hours the feed dosing has been stopped in order to investigate the effect of substrate limitation during the last stage of the cultivation.

The entire AFF1108 cultivation lasted for 203 hours and a total of 50 samples have been withdrawn. 1 hour sample intervals have been used in the periods 6 to 22h and 31 to 41h. In the period from 45 to 198h a sampling interval of approximately 8 hours has been used. Thereafter samples have been taken at 201, 202 and 203 hours.

Each sample is stored in a room having a temperature of 4°C immediately after it has been taken. When an appropriate number of samples has been collected the analyses have been carried out. It was belived that storage at this temperature did not affect the subsequent analysis, but an investigation has shown that the measured biomass can decrease by around 30% after 3 days of storage. This behaviour is most likely to be connected with cell lysis. This issue is not treated further in this work but it must be kept in mind that fresh samples give the most reliable biomass measurement.

The types of analyses carried out on the samples are given below, however, not all analyses are carried out on all samples, some analyses are made less frequently.

1. Biomass measurements (Cell dry matter)
2. Enzyme activity (Samples are sent to another facility for analysis)
3. HPLC analysis (Glucose and metabolites)
4. Particle size measurement
5. Microscopy images

Detailed information concerning the experimental methods and results are given in appendix B.

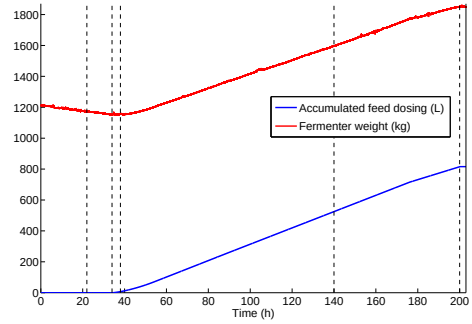
6.3 Experimental results

6.3.1 Overview of cultivation

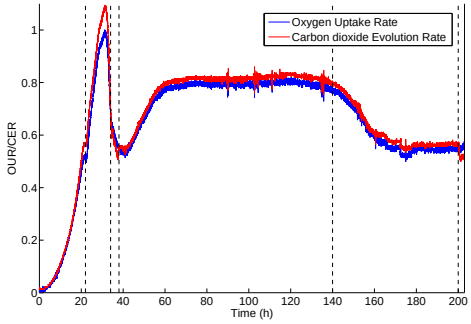
The evolution of the cultivation is similar to the ones run by a standard recipe [1] up until around 150 hours where the DOT drops to around 3%. The feed dosing rate has been decreased at this point to avoid a more severe oxygen limitation. The normal procedure would be to decrease the feed dosing rate at an earlier stage but in order to investigate the metabolism under oxygen limitation the correction has been postponed. It has been found that the evolution of the cultivation can

[1] Appendix A contains information on the evolution of a standard batch

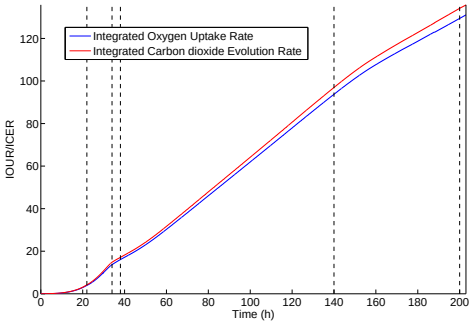
be divided into distinct operating regions based on important physiological factors. Six such operating regions have been identified and are described in the following. The samples have been analysed for glucose and metabolites by HPLC (Appendix B). A number of compounds have been identified but only glucose, malate and mannitol were found to have significant concentrations. An overview of the most important cultivation variables is given in figure 6.1 and 6.2. The six operating regions mentioned are indicated in the plots by dashed vertical lines.



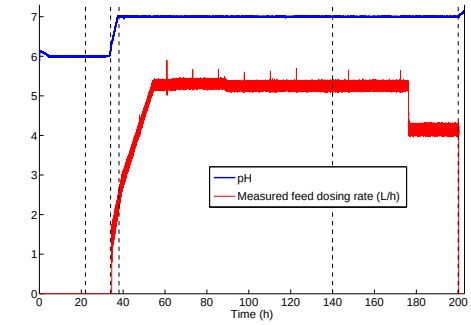
(a) Accumulated feed dosing (L) and broth weight (kg)



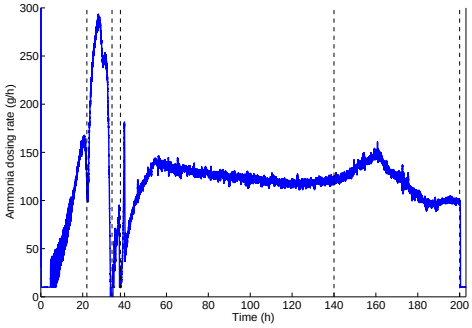
(b) Scaled OUR and CER (moles/h)



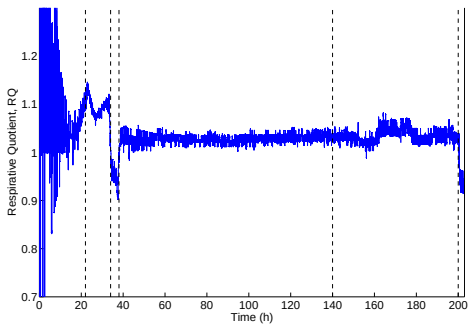
(c) Scaled Integrated OUR and Integrated CER (moles)



(d) pH and measured dosing feed rate (L/h)

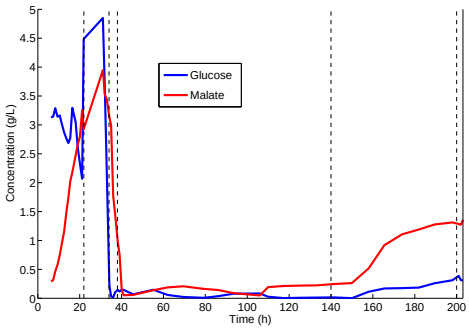


(e) Ammonia dosing rate (g/h)

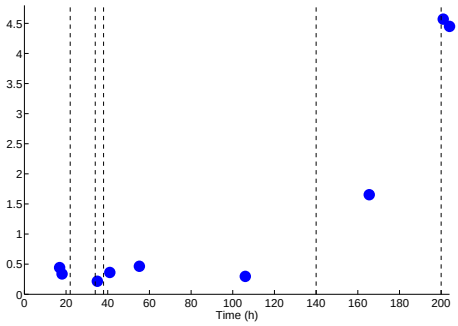


(f) RQ

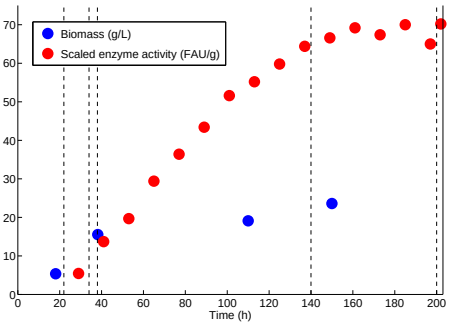
Figure 6.1. Overview of AFF1108 fermentation. The division into operating regions is indicated by vertical dotted lines



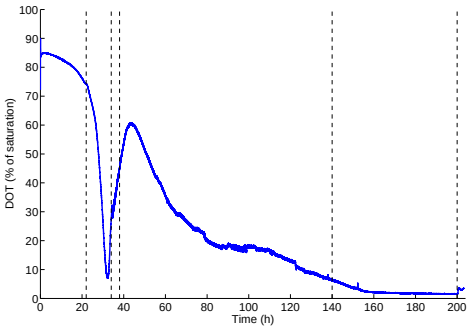
(a) Glucose and malate concentration (g/L)



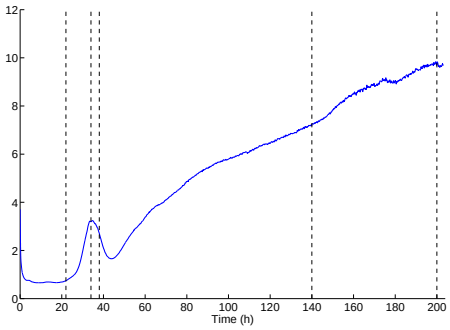
(b) Mannitol concentration (g/L)



(c) Biomass and enzyme activity



(d) DOT



(e) Viscosity

Figure 6.2. Overview of AFF1108 fermentation. The division into operating regions is indicated by vertical dotted lines. In subfigure (a) a linear interpolation has been made between the measured points to improve clarity

0-22h (Period 1a): Rapid growth during batch phase

This period is characterised by a rapid biomass growth rate (figure 6.2 (c)) due to the high substrate concentration. The OUR and CER increase rapidly but show a small stagnation at 21h (figure 6.1 (b)). The RQ lies around 1.03 in the beginning and rises to above 1.1. The pH is 6.0 during this and the following period and the feed dosing is not started before 34h. The DOT starts at 85% and decreases to 74% at 22h. The ammonia dosing rate increases rapidly due to the fast growth but makes a sudden drop just before 22h, which coincides with the observed stagnation of OUR

and CER. The glucose concentration seems to be at a relatively high level between 2 and 5g/L up to 34h (figure 6.2 (a)). The reason why the glucose concentration is lower than the total substrate concentration (50g/L) is that the initial substrate is supplied as starch which is hydrolysed into glucose by the enzymes produced by the organism and this process takes some time. An explanation for the large fluctuations in the measured glucose concentration could be that the degradation of starch and metabolism of glucose happen simultaneously and if this happens at different rates no constant concentration of glucose is observed. A very large malate formation is observed during this period, the concentration reaches around 3g/L at the end of the period. This is a very significant concentration compared to the biomass which reaches a concentation of almost 5.5g/L at 18h. Relatively little product is formed in this period and the mannitol concentration stays below 0.5g/L.

22-34h (Period 1b): Formation of metabolites during batch phase

In this period the OUR and CER again increase rapidly. They peak at 32h and decrease towards the end of the period. This decrease coiincides with a decrease in the glucose concentration which is practically depleted at the end of the period. The DOT decreases to 7% at 32h and then increases rapidly to 27% at 34h. The RQ stays at a high level of around 1.1. A small amount of malate and almost no mannitol is produced. The ammonia dosing rate rises from the sudden drop just before 22h and peaks at 27h and then decreases rapidly to zero at the end of the period. The second operating region is characterised by very large and fast changes. The experimental results suggest that the metabolic activity first peaks and then decreases drastically as substrate limitation sets in.

34-39h (Period 2): Metabolite consumption

This period only covers 5 hours of the cultivation but captures a fundamental change in the metabolism. The substrate is depleted but the concentration of malate decreases to a level close to zero while the mannitol concentration is unchanged. The feed is started at 34h at a low rate meaning that a certain concentration of substrate must be present but apparently it is not large enough to cover the carbon and/or energy demand of the micro-organism since malate is continuing to be consumed. Practically all the malate formed during the batch period is metabolised during this 5 hour period which also explains why OUR and CER only decrease slightly and do not drop to zero as would be expected if glucose were the only carbon and energy source. As a consequence of the decreased oxygen demand the DOT increases to 45% during the period. RQ drops significantly to 0.94. Changes in RQ are most often connected to changes in the metabolism which supports the assumption of change in substrate. pH increases spontaneously and the pH setpoint for the controller is increased gradually during this period from 6.0 to 7.0. This higher set-point is kept for the remainder of the cultivation.

39-140h (Period 3): Substrate limitation leading to large production rate

This period covers the most productive part of the cultivation. Production of enzyme is enhanced by glucose limitation and the measurements show that the glucose

concentration in fact is very low during this period. The enzyme activity (figure 6.2 (c)) increases almost linearly except for the last part of the period, indicating a nearly constant production rate. The growth rate of biomass is very low which is a consequence of the low glucose concentration. OUR and CER stabilise during the first 20 hours of the period and remain at a constant level up to 140h. The DOT reaches a peak of almost 61% at 43h and decreases thereafter. It seems to stabilise at 17% at 100h but then starts a new decrease and ends at 6.5% at 140h. The feed dosing rate is too high to maintain a constant DOT level thus under normal operation the feed dosing rate would have been reduced by the process operators around 110 to 120h. Both the malate and mannitol concentrations stay almost constant and the RQ stays around 1.03.

140-200h (Period 4): Oxygen limitated period

During this period the effects of oxygen limitation start to appear. The DOT decreases at the beginning of the period and stabilises at around 1.5% between 170 and 200h which is a very low value for this cultivation. At 176h the feed dosing rate is decreased by 20% to avoid too severe consequences of the oxygen limitation. OUR and CER drop and glucose starts to accumulate in the cultivation tank. The cultivation has now switched from being limited by substrate to be limited by oxygen. A large production of mannitol is observed and also a certain amount of malate is formed. The production rate of enzyme decreases drastically which could be due to a combination of substrate repression and oxygen limitation.

200-203h (Period 5): Metabolite consumption

In the last few hours of the cultivation the feed dosing has been stopped and the effect on DOT is seen almost immediately. It increases rapidly from 1.5% to 3.5% during 30 minutes. The value then drops a little bit to 3% at 202h and increases slowly to 3.5%. OUR is not affected very much whereas CER drops significantly leading to a drop in RQ from 1.03 to a little above 0.9. These observations suggest that the metabolism changes radically most likely due to glucose limitation. The pH rises shortly after the feed dosing is stopped. This was also observed at the shift between batch and fed-batch and could indicate that glucose has become limiting.

An interesting observation made above is that DOT increases even though OUR is constant or even increasing. This observation can be explained by hyphal fragmentation. Fragmentation of hyphae should give a reduced viscosity leading to a better oxygen transfer. Unfortunately this hypothesis can not be supported by the online measurement of viscosity as the signal from the viscosimeter towards the end of the cultivation is too noisy (even after filtering) to permit observation of even significant changes.

6.3.2 Summary of observations

The cultivation is started with a high initial amount of substrate leading to rapid growth of biomass. As soon as the initial substrate is depleted the biomass growth practically stops. Feed dosing is started and the feeding strategy ensures low substrate concentration and gives a much lower growth rate combined with a

large production rate. The rapid growth in the batch phase is connected with a high RQ and fairly high production rate of malate. Production of malate is also observed after 150h where oxygen limitation sets in and the substrate concentration increases. This suggests that malate formation is a consequence of high substrate concentrations. The ammonia feed dosing rate can be seen as an indicator of the combined production rate of biomass and enzyme. A sudden drop in ammonia feed dosing just before 22h indicates a drop in the biomass formation at this point. After this point the RQ reaches a high value of 1.1. When the substrate is exhausted the organism adapts to the new conditions and begins to metabolise malate. This phenomenon is observed both at feed start and at the end of the batch when the feed dosing was shut down for 3 hours. The severe oxygen limitation starting at around 150h leads to production of mannitol. This is the only period where such a large mannitol production is found.

An overview of the identified operating regions is given in table 6.1.

Table 6.1. Overview of the seven operating regions identified in section 6.3.1

Operating region	Time period (h)	Duration (h)
1a Rapid growth during batch phase	0-22	22
1b Formation of metabolites during batch phase	22-34	12
2 Metabolite consumption	34-39	5
3 Substrate limitation leading to large production rate	39-140	101
4 Oxygen limited period	140-200	60
5 Metabolite consumption	200-203	3

6.3.3 Mechanisms for malate formation

The formation of malate during high glucose concentrations can be explained by a limitation in the citric acid (TCA) cycle. The TCA cycle plays a very important role in cell metabolism and is a part of the mechanism converting carbon substrates to carbon dioxide (Madigan *et al.* (2000)). Malate is a 4-carbon compound appearing in the TCA cycle, the chemical formula is $C_4H_6O_5$. A higher carbon substrate uptake in the cell results in a higher flow through the TCA cycle and it is therefore reasonable to assume that the capacity in this cycle may not be sufficient to handle the high substrate uptake rates.

6.3.4 Mannitol formation during oxygen limitation

Mannitol formation is only observed during periods of very low oxygen concentration. An obvious thought is therefore to investigate if mannitol production can be connected to a mechanism present during oxygen limitation. The cell needs some oxidising compound to oxidise $NADH$ to NAD^+. NAD is short for nicotinamide adenine dinucleotide, it serves as a means of carrying reduction equivalents in the

cell. It exists in an oxidised form as NAD^+ (degree of reduction of zero) and a reduced form as NADH (degree of reduction of 2). When sufficient oxygen is present the oxidation of NADH occurs by oxidative phosphorylation in respiration. The equation for respiration is often written in a short form as given below:

$$-NADH - \frac{1}{2}O_2 = 0 \tag{6.1}$$

Species like protons, water, ADP, ATP and NAD^+ are also involved in respiration but for simplicity are often not considered in the overall reaction. NADH has a degree of reduction of 2 and each oxygen atom has -2 which balances the equation. Another mechanism for reducing NADH is by substrate-level phosphorylation in the fermentation reaction. In that reaction a substrate is reduced instead of oxygen as in respiration.

Mannitol ($C_6H_{14}O_6$) has a degree of reduction per carbon of 4.33 compared to 4 for the substrate, which means that mannitol could in fact be product of a substrate-level phosphorylation. In this case the respiration in equation 6.1 would be substituted by the fermentation equation given below:

$$-\frac{1}{6}NADH - CH_2O + CH_{7/3}O = 0 \tag{6.2}$$

This fermentative reaction produces much less ATP than the respiration (Madigan *et al.* (2000)). ATP (Adenosine Triphosphate) consists of adenosine, to which 3 phosphate molecules are bound in series. ATP serves as the most important energy carrier in the cell and is generated in exergonic reactions (reactions with negative Gibbs free energy) and is being used to drive endergonic reactions. The substrate is used more efficiently for producing energy in respiration than in the fermentation reaction. This also means that much lower yields of products can be expected during periods of oxygen limitation.

6.4 Balances

This section deals with analysing three important balances, namely carbon, nitrogen and reduction balances. The balances provide an indication of the quality of the experimental data and also whether all major species have been identified. The composition and degree of reduction of the constituents considered in the balances are given in table 6.2. For compounds containing carbon the composition and degree of reduction is reported per carbon mole. The table covers all the compounds identified in significant amounts.

The composition of biomass and product are taken from Carlsen (1994). The strain in Carlsen (1994) is a predecessor to the one used in this work and it is assumed that the same composition is valid.

The substrate composition is written as $CH_{2x}O_x$ in order to account for the substrate being composed of monomer chains. Each time a monomer is added to a chain one molecule of water is released. The ratio between H and O will therefore always be 2:1. This influences the molar weight but not the degree of reduction. Appendix A.3.1 gives the calculation of actual molar weight of the feed dosing. All the balances have been set up by considering periods of the cultivation. Balances

Table 6.2. Composition and degree of reduction of the compounds included in the balances. Substrate accounts for both initial substrate and added feed dosing

Compound	Composition	Degree of reduction, κ
Biomass/C mole*	$CH_{1.72}O_{0.55}N_{0.17}$	4.11
Product/C mole	$CH_{1.50}O_{0.31}N_{0.25}S_{0.008}$	4.15
Substrate/C mole	$CH_{2x}O_x$	4
Mannitol/C mole	$CH_{7/3}O$	4.33
Malate/C mole	$CH_{3/2}O_{5/4}$	3
Oxygen/mole	O_2	-4
Carbon dioxide/mole	CO_2	0
Ammonia/mole	NH_3	0

* An ash content of 7.5w/w% is asssumed.

for each constituent have been made as changes of moles of that constituent in the given period. Also total balances have been formulated covering the entire duration of the cultivation.

6.4.1 Carbon balance

The carbon balance has been set up considering the carbon containing compounds in table 6.2. The initial substrate and the added dosing are both corn starch and it is assumed that the content of carbon which can be metabolised is the same in both cases. The calculation of the carbon concentration is given in appendix A.3.1. Throughout this work the density of the broth is assumed to be 1.03kg/L. The amount of substrate which is transferred from the seed tank is assumed negligible as the microorganisms have consumed almost all the substrate in the seed tank at the point in time where the transfer is made.

The balance is calculated at the 4 time points where biomass measurements made on fresh samples are available. It is assumed that the amount of biomass at the beginning of the batch is negligible. The measured enzyme activities have been linearly interpolated to coincide with the points of measured biomass concentration. The four periods used in the carbon balance are not identical to the operating periods mentioned in 6.3.1.

Table 6.3 gives the results of the carbon balance. The overall recovery is 89.3% which is not impressive compared to laboratory scale experiments. However for pilot plant scale it is considered reasonable. During the first period between 0 and 38.17h the recovery is below the overall recovery. The second period which goes up to 110h is above the overall recovery. Thus one or more compounds may be missing during the first period and is metabolised during the second period. In this second period more carbon is utilised than added by the actual dosing giving a higher apparant recovery. The third period is part of the most productive operating region and is not expected to have any significant metabolite formation. In fact the recovery during this period (90.4%) is very close the overall recovery (89.3%). The fourth period has

Table 6.3. Carbon moles added and recovered during each of the 4 periods

Period	0-38.17h	38.17-110h	110-150h	150-202h	0-202h
Biomass	224.3	593.5	314.1	441.3	1573.1
Product*	64.9	313.1	145.8	95.5	619.4
Carbon dioxide	confidential	confidential	confidential	confidential	confidential
Glucose	5.10	-3.92	-1.10	17.6	17.6
Malate	34.4	-25.9	4.56	58.17	71.3
Mannitol	10.3	6.88	47.6	189.1	253.9
Recovered	1457.9	4893.9	2708.2	2742.8	11803.0
Added	1725.7	5105.0	2995.7	3387.8	13214.3
Recovery	84.5%	95.9%	90.4%	81.0%	89.3%

*Due to confidentiality the enzyme yields are scaled by a constant factor

a very low recovery of only 81.0% and this might very well be due to unidentified metabolites being formed during this oxygen limited period.

Table 6.4. Carbon recovered in each of the 4 periods, calculated as mole fractions of added carbon

Period	0-38.17h	38.17-110h	110-150h	150-202h	0-202h
Biomass	0.130	0.116	0.105	0.130	0.119
Product*	0.0376	0.0614	0.0486	0.0282	0.0468
Carbon dioxide	confidential	confidential	confidential	confidential	confidential
Glucose	0.00295	-0.000767	-0.000367	0.00518	0.00133
Malate	0.0199	-0.00507	0.00152	0.0172	0.00539
Mannitol	0.00596	0.00135	0.0159	0.0558	0.0192
C-balance	0.845	0.959	0.904	0.810	0.893

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table 6.4 gives the results of the carbon balance expressed as mole fractions of the added carbon. These can be regarded as overall yield coefficients for conversion of carbon substrate into the respective compounds. It is interesting to observe that the yield coefficient for conversion of substrate into carbon dioxide have relatively small variations (0.50 ± 0.04). In the first time period the yield coefficient for biomass is relatively high and the yield coefficient for product is relatively low. This period covers the batch phase and the result is expected because of the high substrate concentration. The second period has the largest yield coefficient for product formation which fits well with it being the period having the most optimal conditions for product growth, namely low substrate concentration and sufficiently high DOT. During the third and especially fourth period the yield of product decreases drastically which is also expected from the combination of substrate repression and oxygen limitation. The yield of biomass during the fourth period is as large as in the first period (the

batch phase) which supports the hypothesis of morphological changes in filamentous fungi during oxygen limitation as observed by Zangirolami (1998). Rapid growth of filaments during oxygen limitation were observed by Zangirolami (1998) leading to increasing viscosity and further impaired oxygen transfer. In fact this type of response may be described as a catastrophic response ultimately leading to death of biomass if oxygen limitation prevails. As expected the yield coefficient for malate is relatively high in the first period and changes to be negative in the second period indicating that malate is being consumed. The largest yield coefficient for mannitol is found in the fourth period where oxygen limitation is prevailing.

The overall carbon balance fits relatively well considering the circumstances but one or more unidentified compounds are likely to be present.

6.4.2 Nitrogen balance

A nitrogen balance has been formulated including the nitrogen containing compounds:

1. Initial nitrogen from added salts
2. Added gaseous ammonia
3. Recovered biomass
4. Recovered enzyme
5. Measured ammonia concentration in the cultivation tank

The initial nitrogen comes from ammonium sulfate.The amount of nitrogen transferred from the seed tank is assumed to be negligible. It is assumed that no nitrogen leaves the cultivation tank, that no nitrogen is present in the dosing medium and that the nitrogen from the air is inert. The pH value around 6 and 7 means that almost all the ammonia is dissolved as ammonium ions in the cultivation broth.

The nitrogen balance is calculated at 7 time points as shown in table 6.5 where the points have been chosen to correspond to off-line measurements of ammonia concentration and enzyme activity. The 4 biomass points used in the carbon balance have been linearly interpolated to correspond to these 7 time points.

The overall recovery is 88.3% which is almost identical to the carbon recovery (89.3%). Again differences between the periods are observed. The three first periods which cover up to 101h all have recoveries above the overall recovery. The four last periods all have recoveries lower than the overall one and the recoveries show a decreasing tendency. This suggests that either a nitrogen containing species is not identified or that one or more of the measurements are not reliable.

Table 6.6 gives the results of the nitrogen balance expressed as mole fractions of the added nitrogen. These can be regarded as overall yield coefficients for conversion of supplied nitrogen into the respective compounds. An interesting observation is that the distribution of nitrogen between biomass and enzyme product changes significantly during the cultivation. Also there is a large variation in how much ammonia actually stays dissolved in the cultivation tank.

Table 6.5. Changes in moles of nitrogen added in each of the 7 periods and for the entire cultivation

Period	0-53h	53-77h	77-101h	101-125h	125-149h	149-173h	173-202h	0-202h
Biomass	126.61	21.98	23.83	34.29	43.44	32.49	32.64	315.3
Product*	28.08	28.72	30.68	21.60	21.06	11.06	13.74	154.92
Broth	conf.	conf.	conf.	conf.	conf.	conf.	conf.	conf.
Total rec.	328.5	176.0	180.6	142.1	146.8	135.3	121.0	1230.2
Total add.	334.4	187.8	176.0	167.9	170.9	191.4	164.4	1392.8
Recovery	98.2%	93.7%	102.6%	84.6%	85.9%	70.7%	73.6%	88.3%

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table 6.6. Nitrogen recovered calculated as mole fractions

Period	0-53h	53-77h	77-101h	101-125h	125-149h	149-173h	173-202h	0-202h
Biomass	0.379	0.117	0.135	0.204	0.254	0.170	0.199	0.226
Product*	0.084	0.153	0.174	0.123	0.123	0.058	0.084	0.111
Broth	conf.	conf.	conf.	conf.	conf.	conf.	conf.	conf.
Balance	0.982	0.937	1.026	0.846	0.859	0.707	0.736	0.883

*Due to confidentiality the enzyme yields are scaled by a constant factor

6.4.3 Degree of reduction balance

A degree of reduction balance has been formulated considering all the species having a degree of reduction different from zero. Ammonia is not taken into account as it has a degree of reduction of zero (Nielsen *et al.* (2003)). The components transferred from the seed tank are also assumed negligible in this balance. Each term in the balance is calculated as the degree of reduction per mole multiplied by the number of moles of the particular compound.

The degree of reduction balance is based on the following overall reaction:

$$biomass + products + CO_2 - substrate - NH_3 - O_2 = 0 \tag{6.3}$$

The reduction balance is written as:

$$balance = \kappa_{biomass} + \kappa_{products} - \kappa_{O2} - \kappa_{substrate} \tag{6.4}$$

If all the compounds are identified correctly the balance should equal zero. As for the carbon balance 4 different points corresponding to measurements of biomass have been used. The enzyme activity points are interpolated linearly to correspond to these time points. The results are given in table 6.7.

Overall fewer reduction equivalents are recovered than added. The balance actually fits relatively poorly compared to the carbon and nitrogen balances. It must be stated though that the degree of reduction balance is very sensitive to uncertainties in measurements of substrate feed and oxygen. This is because the number of degree of reduction equivalents added is the sum of the contributions from oxygen and substrate. These have opposite sign meaning that small errors in these measurements

Table 6.7. Degree of reduction equivalents added and recovered in each of the 4 periods and for the whole cultivation

Period	0-38.17h	38.17-110h	110-150h	150-202h	0-202h
Biomass	922	2439	1291	1817	6469
Product*	270	1299	605	398	2572
Glucose	20.4	-15.7	-4.4	70.3	70.6
Malate	98.6	-74.1	13.1	167	204
Mannitol	47.0	31.4	217	865	1160
O_2	conf.	conf.	conf.	conf.	conf.
Feed	6903	20420	11982	13551	52857
Recovered	2435	8878	4542	4908	20763
Added	3674	9669	5731	7687	26761
Recovery	66.3%	91.8%	79.3%	63.9%	77.6%

*Due to confidentiality the enzyme yields are scaled by a constant factor

can be magnified when the ratio between recovered and added degree of reduction equivalents is calculated.

An interesting observation is that the period with highest productivity (the second period) shows the highest recovery (91.8%). This is also expected to be the period where the least amount of metabolic byproducts is produced. The first and the fourth period have the lowest recoveries which is the same picture as seen for the carbon balance. This indicates that one or more compounds with a degree of reduction larger than zero have not been identified. As for the carbon balance the reason for the very high recovery during the second period might be that unidentified metabolites formed during the batch phase are consumed.

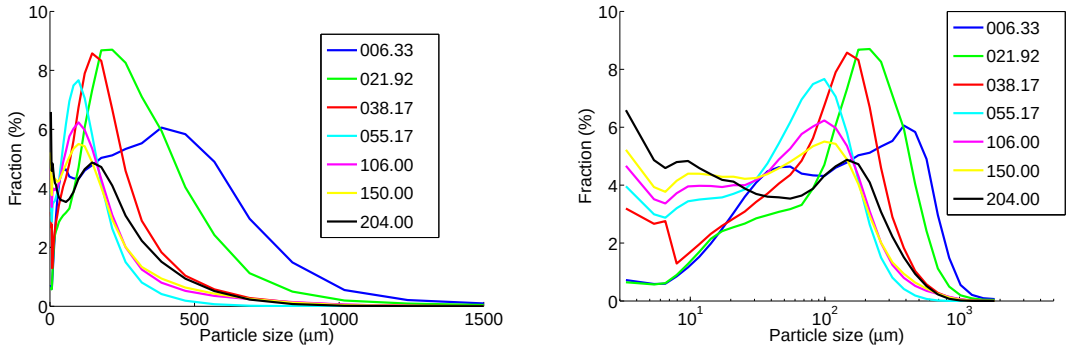
6.5 Morphology

This section provides a study of the progress of morphology during the AFF1108 cultivation experiment. Particle size distribution measurements have been carried out on the samples and are combined with microscopical images to provide an understanding of the morphology in an attempt to reveal how this may relate to product yield.

6.5.1 Particle size distribution

Particle size distributions have been measured at 32 time points during the cultivation. All the raw measurements are given in appendix C.1. The measurements have been performed on a Malvern Mastersizer X instrument connected to a computer running the Mastersizer X v3.10 software. Particle sizes between 1.8 and 2000μm are measured and divided into 32 logarithmic distributed size classes. The distributions are calculated on a volume basis. The apparatus is not originally intended for making measurements on fungi, but rather on suspensions of spherical particles. The obtained measurements should therefore be used with great care. Experience

shows that the obtained particle volume can be regarded as a corresponding volume of the pellet or agglomerate of hyphae (Stocks (2005)). Despite the uncertainties of the measurements the results can provide an indication of the macroscopic morphology present throughout the duration of the cultivation. The distinction between freely dispersed hyphae and pellets is very difficult to make. For *Aspergillus oryzae* it is often assumed that particles above $300\mu m$ can be regarded as pellets (Stocks (2005)). The pellet size depends heavily on growth conditions and on pellet density, since pellets can have a more or less dense structure.



(a) Linear particle size axis

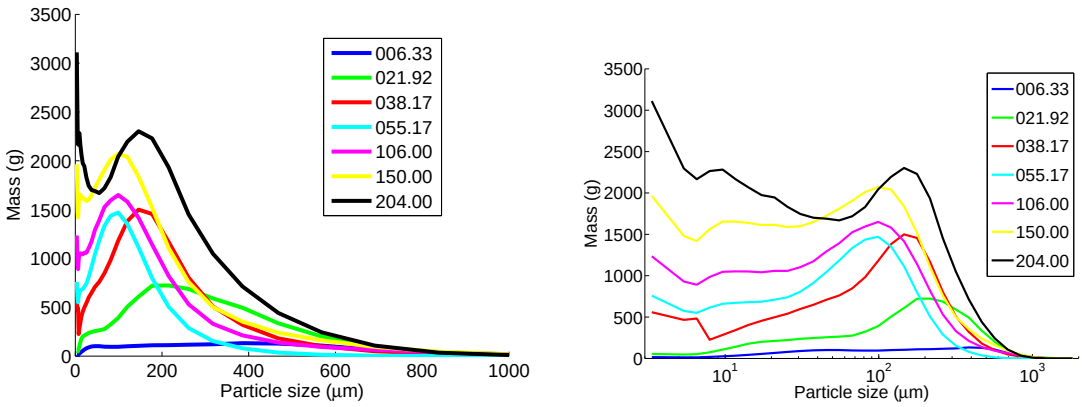
(b) Logarithmic particle size axis

Figure 6.3. Particle size distributions for the AFF1108 experiment. X-axis: Particle diameters in μm. Y-axis: Distribution in percent

Figure 6.3 presents a selection of particle size distributions which cover the evolution throughout the cultivation shown at seven selected time points. The distributions are shown with both a linear and a logarithmic particle size axis. The logarithmic plot enhances the resolution for the smaller particle sizes and it also provides equal spacing between size classes on the particle size axis. A maximum in the distribution is observed which moves as the cultivation progresses. The early stage of the cultivation (6.33h) is dominated by large particles, here the peak is around $400\mu m$. The particle sizes decrease as the cultivation proceeds which makes the maximal distribution move down to around $100\mu m$ at 55-106h. At the very end of the cultivation (150-204h) the maximum moves up to around $150\mu m$. Another interesting observation is that the fraction of very small particles with a diameter of $30\mu m$ or less stays low during the batch part (first 34h) of the cultivation but grows significantly during the fed-batch part. Average diameters as well as other raw and calculated data for all measured time points are given in appendix C.1.

6.5.2 Distribution of biomass

In this subsection the particle size distribution (psd) is multiplied by the total amount of biomass to give a clearer description of the biomass distribution. Provided that the biomass density is constant the volume based psd can also be regarded as a mass distribution. The density is not necessarily constant for dense pellets, so in the current investigation it should be kept in mind that dense pellets could contain more biomass than their volume suggests.



(a) Normal plot (b) Logarithmic plot

Figure 6.4. Distribution of total biomass. X-axis: Particle diameters in μm. Y-axis: Mass distribution in g.

Figure 6.4 shows the particle size distribution multiplied by the total amount of biomass for the above selection of seven time points. Again both linear and logarithmic particle size axes are given. The picture is quite similar to the one seen in the previous plots but with some important differences. The psd plots in figure 6.3 show that the relative volume of sizes above $500\mu m$ decreases but when the actual mass is considered it is seen that the mass of particles in this size class actually increases. In fact the mass within all size classes below $300\mu m$ increases, but during the fed-batch phase the growth of particles below around $50\mu m$ is significantly larger than for the other size classes. Many of these small particles are likely to be fragments from the rest of the biomass. The particle size distribution, however, does not reveal if these fragments are actively growing cell material or dead fragments from e.g. cell lysis. In the following discussion the former possibility is assumed.

6.5.2.1 Time evolution of biomass distribution

Figure 6.5 shows the time evolution of the total amount of biomass within the four size groups indicated. The evolution of the total mass of particles within each of the 32 original size groups is divided into these four groups in order to reflect distinct observed characteristica. The mass of very small particles between 1.8 and $11\mu m$ increases suddenly just around the feed start which is most likely connected to fragmentation. The next group between 11 and $108\mu m$ which is expected to consist mostly of free hyphae increases constantly throughout the duration of the cultivation and exhibits no sudden jumps. The third class between 108 and $286\mu m$ decreases its mass after the feed start and the mass begins to increase at around 55 hours. The largest group (286 to $2000\mu m$) is expected to mostly contain pellets and its mass increases in the very beginning but starts to decrease at around 10 hours and begins to increase again at around 55 hours. It is very interesting to note that the mass of these large pellets is only slightly larger at the very end of the cultivation than at 10 hours even though the total biomass has increased significantly. There is no reason to believe that no large pellets are produced between 10 and 55 hours so

the apparent explanation is that a balance exists between pellet growth and shear forces breaking up the pellets. Pellets grow larger as the organism grows and the decrease in mass of the large pellets between 10 and 55 hours may be interpreted as a slower growth of these. Rapid growth of large pellets after 55 hours is observed and around 150 hours the growth is especially rapid. This coincides with point at which oxygen limitation starts. Zangirolami (1998) observed a rapid growth of hyphae of *Aspergillus oryzae* during oxygen limitation. In this work free hyphae branching out from the center of the biomass were observed. This observation could indicate that particles in the size class ranging from 286 to 2000μm, which have so far been considered pellets, could very well include large agglomerates of freely dispersed hyphae

These findings lead to the hypothesis that particles larger than about 100μm break up due to substrate limitation and produce fragments of sizes less than 11μm.

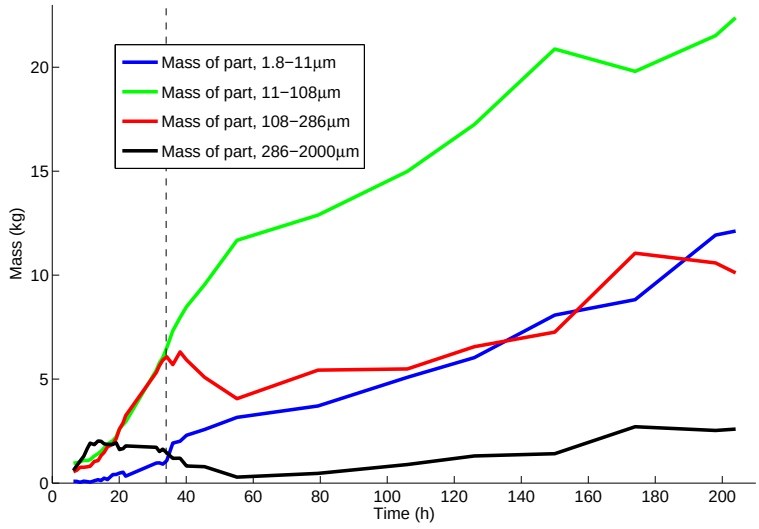


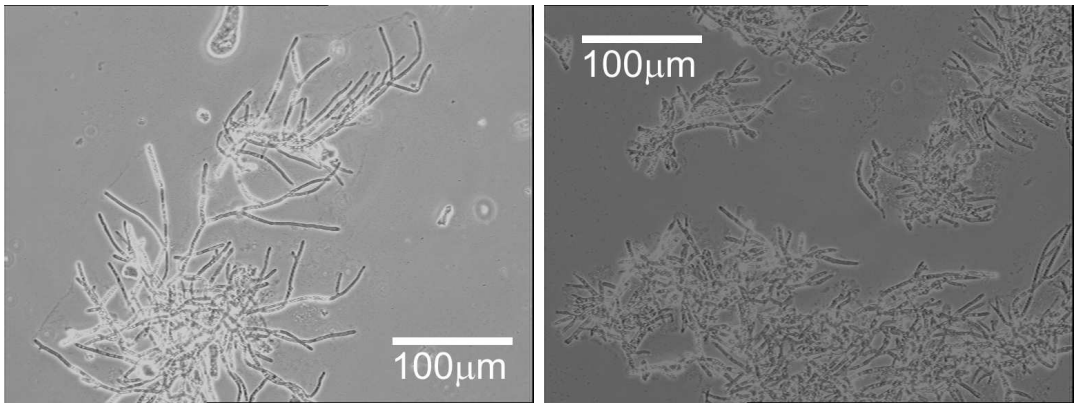
Figure 6.5. Time evolution of total mass within the 4 size categories. The dashed vertical line indicates the dosing feed start

6.5.3 Analysis of microscopy images

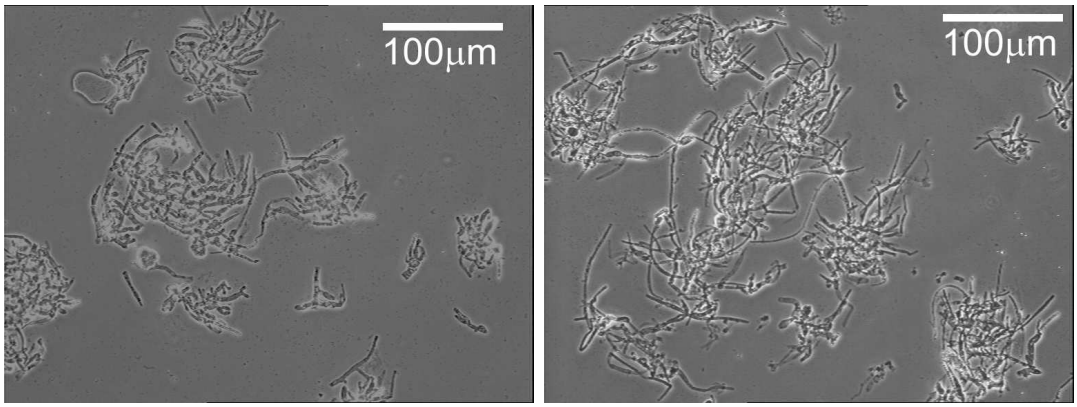
Microscopic images can potentially provide important information on the morphology and structure of the biomass but must also be treated with great care as there is no guarantee that representative information is obtained. In reality it is very difficult to collect a truly representative sample for microscopy analysis and it is therefore very difficult to estimate how the biomass is distributed between different morphologies.

Images from a microscope using 10x, 40x and 100x objective lenses have been studied. Four illustrative images are shown in figure 6.6. More images are given in appendix D.

It can be seen on figure 6.6 that the pellets are definitely larger in the batch phase (figure 6.6 a) than in the fed-batch phase (figure 6.6 c and d).



(a) 8.33 hours (b) 35 hours



(c) 55.08 hours (d) 133 hours

Figure 6.6. Images from microscope with a 10x objective lens

Closer inspection of the images in appendix D reveals that the batch phase is characterised by large pellets and very few freely dispersed hyphae. The pellets can best be characterised as smooth pellets as described in section 2.5.1.1. The images do not reveal whether the pellets are hollow or not. In the fed-batch phase (figure 6.6 c and d) the pellets are smaller and the number of freely dispersed hyphae, which possibly are fragments from pellets, is found to be large. Towards the end of the batch (figure 6.6 d) the cultivation has been oxygen limited and this is reflected in the images. The structure of the free hyphae changes and becomes more widespread.

6.5.4 Division into different morphologies

The particle size distribution analysis combined with images of broth samples indicates that the very beginning of the cultivation is dominated by large pellets. After the feed has started the cultivation is dominated by smaller pellets and freely dispersed hyphae.

The formation of freely dispersed hyphae can have several explanations. The major differences in growth conditions between the batch and fed-batch phase are pH and substrate concentration. It is well known that pH has a large effect on pellet formation. Carlsen (1994) found that lower pH values favour the freely dispersed

morphology. Fragmentation of large pellets also contributes to the amount of freely dispersed hyphae.

The shear rate caused by the stirring also has a large effect on the pellet size (Li *et al.* (2002)). The stirring speed has been kept constant throughout the batch except towards the very end where it has been decreased due to high viscosity. The pellets grow in size as the microorganism grows but the shear rate causes the pellets to break when they become too large. After a certain period of time an equilibrium size can be expected. The pellets transferred from the seed cultivation tank to the main cultivation tank seem to be larger than this equilibrium size.

Li *et al.* (2000) found that the size of the hyphal elements decreased due to the stirring intensity. When the shear force is larger than the tensile strength of the cell wall fragmentation will occur. At some point in the cultivation a balance is obtained and the pellet size remains more or less constant.

From the results obtained it is not straightforward to distinguish pellets from free hyphae. In fact the psd could be interpreted as two distributions, one for each kind of morphology, overlapping each other and producing a bimodal distribution. To help distinguish between free hyphae and pellets figure 6.7 is shown. Here the mass fractions of particles above 5 different particle sizes are plotted as function of time.

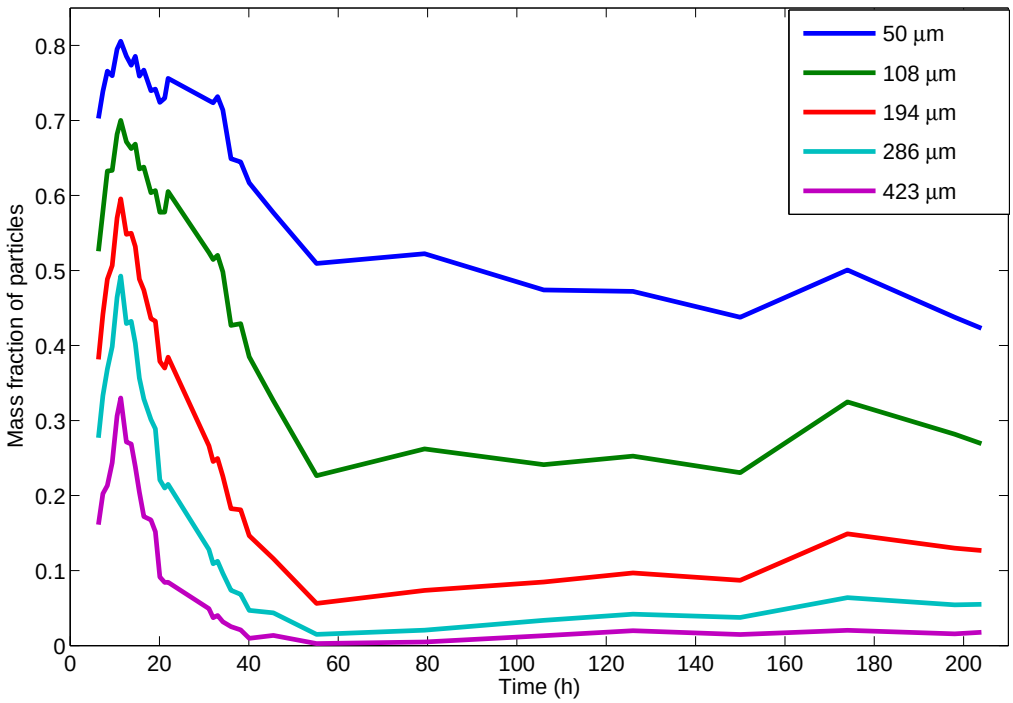


Figure 6.7. Mass fraction of particles larger than size indicated in the legend. Five different size divisions are shown. The figure shows that the massfraction of large particles decreases dramatically as the batch phase stops. The batch phase stops at 34h

Figure 6.7 shows that no matter at which particle size the division between free hyphae and pellets is made the fraction of pellets in the batch phase is very large

compared to the fed-batch phase. Using a division at $50\mu m$ more than 70% of the mass is pellets in the batch phase dropping to around 50% in the fed-batch phase. Using a division at $423\mu m$ the mass fraction of pellets reaches a maximum of around 32% in the batch phase and drops to very few percent in the fed-batch phase. The figure supports the hypothesis that large pellets are found in the beginning of the cultivation and split up when the feed starts. During the fed-batch phase an equilibrium distribution between the two morphologies is obtained. The observed change in distributions may be explained as a combination of substrate limitation and shear forces.

6.5.5 Relating morphology to viscosity

A common hypothesis is that freely dispersed hyphae have a major impact on viscosity and that pellets only have a negligible effect (Poulsen (2005)). It is therefore interesting to investigate if evidence of this hypothesis can be observed in the present data.

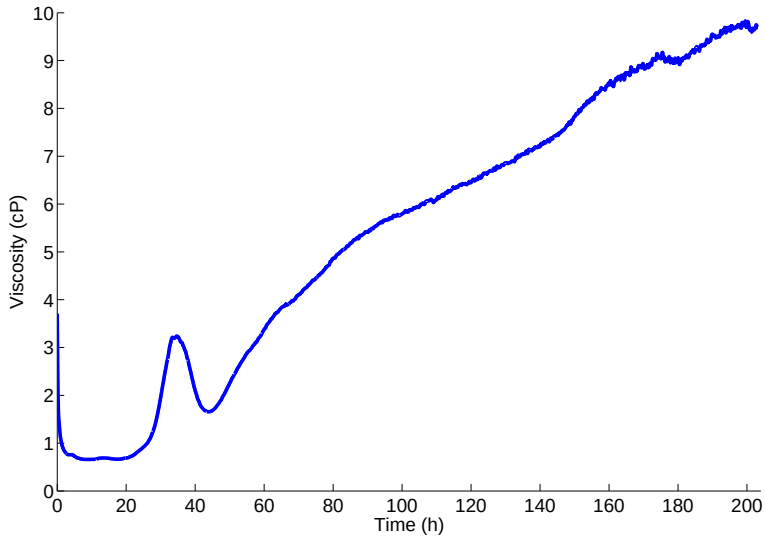


Figure 6.8. Time evolution of the viscosity. The dashed vertical line indicates the dosing feed start

Figure 6.8 shows the evolution in viscosity during the entire batch. In the batch phase mostly pellets are present leading to a low viscosity. Later the distribution changes to lower sizes (figure 6.7) meaning that formation of freely dispersed hyphae occurs. This rises the viscosity significantly. When the substrate is used up starvation occurs and the freely dispersed hyphae split up which reduces the viscosity significantly (figure 6.8). The viscosity then increases almost linearly with time, the size distribution is the same but the amount of biomass increases. Towards the end of the fed-batch phase pellets or large freely dispersed hyphae are formed due to oxygen limitation. This can be seen as a small jump in the size distribution plot (6.7) but it is not observed to contribute to the viscosity. It is noteworthy that the

noise on the viscometer increases during the oxygen limitation period. This could be related to the increased presence of large freely dispersed hyphae.

6.6 Conclusion

This chapter describes the experiment with additional measurements carried out in the Fermentation Pilot plant at Novozymes A/S. The experiment has been operated according to a standard recipe except that the process has been exposed to oxygen limitation towards the end of the cultivation. The major difference from standard monitoring is that extensive laboratory analyses have been carried out and the frequent sampling has been used to identify the trajectories of biomass, enzyme activity and metabolite concentrations.

It was observed that the material balances fit relatively well considering the scale of the equipment. The balances indicated that one or more unidentified compounds, possibly containing nitrogen, are likely to be formed during the batch phase and in the oxygen limited phase. Two metabolites which were not observed before have been identified, namely malate and mannitol. Malate is formed during periods with high glucose concentration which is thought to be a result of an overflow mechanism. Also it was observed that malate can substitute glucose during periods with insufficient glucose supply. Mannitol is formed under oxygen limitation as a product of substrate level phosphorylation but it has not been observed that this compound has been metabolised. The highest conversion rate of substrate into product was obtained at low glucose concentrations and a DOT level above around 6%.

The particle size analysis also revealed a number of important findings. Large pellets are transferred from the seed tank and later these large pellets are broken up due to substrate limitation and shear stress. The fragments produced have very small particle sizes and some time after the dosing feed is started a balance between pellet growth and break up develops. It was also observed that large particles, which are likely to be freely dispersed hyphae with long filaments, are formed during oxygen limitation. The change in morphology can also be related to the viscosity - as the number of free hyphae increases the viscosty also inscreases .

The experiment has contributed significantly to the understanding of many aspects of the cultivation. E.g. it has been observed that the cultivation can be divided into operating regions with distinct features. In some of these regions two important metabolites have been observed and it has been uncovered under which circumstances they appear. This provides a basis for development of model of the entire cultivation which can provide insight into how the existing batch operations model can be altered to provide an even better utilisation of the production equipment.

7

Analysis of cultivation operating regions

7.1 Introduction

This chapter provides a more thorough analysis of the pilot plant cultivation experiment described in chapter 6. The purpose of this experiment was to obtain more detailed information on the cultivation process.

A thorough description of the experimental procedure and results is given in chapter 6.

An important observation made was that the cultivation process can be divided into different operating regions with distinct features. A further analysis of the experimental data obtained based on these divisions is the topic of this chapter.

The chapter is structured in the following way: Section 7.2 provides a determination of the specific rates of growth and enzyme production within the operating regions. This obtained information is used in section 7.3 to propose a model for the entire cultivation. The observations made in this chapter lead to a change in the model structure which is presented in 7.4. A conclusion is given in 7.5.

7.2 Estimation of specific rates

This section analyses the specific production rates of both biomass and products formed, including metabolites. The calculations are based on the measurements obtained in the AFF 1108 cultivation described in chapter 6. Specific and total production rates are estimated within defined six operating regions. The rationale behind this is that it has been found that the conditions for growth and enzyme production are approximately constant within the operating regions. This allows for development of simple models applicable to each operating region.

7.2.1 Estimation of specific growth rate from offline biomass measurements

Estimations of specific growth rates have been performed on the off-line biomass measurements obtained in the AFF 1108 cultivation. The gas analysis can only be used to estimate growth rates if it can be assumed that the yield coefficients relating growth to the gas analysis can be assumed constant throughout the cultivation or if the variation of the yield coeffients is known throughout the cultivation. In this cultivation the yield coefficients are neither known nor can they be assumed constant.

Five biomass measurement points are regarded reliable (the biomass measurement has been carried out on fresh samples, appendix B) and these points are used to find the average specific growth rates between the measurement points. As the specific growth rate is assumed constant within the period it can be found as:

$$\frac{dx}{dt} = \mu x \tag{7.1}$$

which upon integration gives:

$$ln\left[\frac{x_{k+1}}{x_k}\right] = \mu(t_{k+1} - t_k) \tag{7.2}$$

x denotes the total number of moles of biomass and not the biomass concentration. The estimated specific growth rates are given in table 7.1

Table 7.1. Estimated mean specific growth rates inbetween off-line biomass measurements

Period (h)	Estimated mean specific growth rate ($10^{-3}h^{-1}$)
18.08-38.17	51.8
38.17-110	6.07
110-150	8.28
150-204	4.18

Table 7.1 shows that the specific growth rate decreases rapidly in the fed-batch part of the cultivation (the dosing feed starts at 34h). In section 6.3 it was found that the period having optimal conditions for enzyme production is between 39 and 140h. The highest specific growth rates for the fed-batch phase are also found here which indicates that these conditions also support biomass growth. Around 140h the cultivation becomes oxygen limited which is also reflected in the specific growth rate which drops significantly from $8.28 \cdot 10^{-3}$h to $4.18 \cdot 10^{-3}$h. Estimation of the specific growth rate for the first period up to 18.08h is not possible as no biomass measurement is available for the beginning of the cultivation. The first measured biomass concentration at 18.08h is subject to some uncertainty as the concentration is still low at this time point.

The time divisions found in table 7.1 are converted to the the time divisions for the operating regions found in chapter 6. This is done by evaluating the trajectory of estimated biomass and calculating the specific growth rate between the periods dictated by the operating regions found in chapter 6.

Table 7.2 shows the estimated specific growth rates within the operating periods found in chapter 6. These specific growth rates have a larger uncertainty than the ones in table 7.1 because interpolated values have been used.

7.2.2 Estimation of production rates within each operating region

This subsection carries out a further analysis of the substrate uptake rate as well as the rate of production of enzyme and metabolites. The estimations are made based

Table 7.2. Estimated mean specific growth rates fitted to the six operating regions

Period (h)	Estimated mean specific growth rate ($10^{-3}h^{-1}$)
0-22	-
22-34	39
34-38	52
38-140	68
140-200	48
200-203	42

on the division into operating regions found in 6. The production rates for each component are calculated based on a simple mass balance covering each operating period. The experimentally determined concentrations are used at the first and last time point of each operating region. If no experimentally determined value is available at that time point a linear interpolation between two adjacent points has been used. All experimentally determined results are given in appendix A, the trajectories for each component are given in figure 6.1.

Table 7.3. Estimated production rates of each identified component within the six operating regions. All units are in C-moles/h for compounds containing carbon and moles/h for compounds not containing carbon

Period (h)	glc uptake	NH_3	O_2 uptake	CO_2	enzyme**	malate	mannitol
0-22	47.2*	4.1	conf.	conf.	1.0	4.5	0.5
22-34	47.2*	12.3	conf.	conf.	2.1	0.5	-0.3
34-38	27.3	2.6	conf.	conf.	3.8	-17.7	0.6
38-140	72.2	7.1	conf.	conf.	4.2	-0.2	0.4
140-200	68.4	7.0	conf.	conf.	2.1	1.0	3.5
200-203	5.4	0.8	conf.	conf.	-0.2	1.1	-0.05

*The glucose uptake rate in these two periods is an average for the time period 0-34h as reliable measurements of substrate concentration could not be obtained during this period

**Due to confidentiality the enzyme yields are scaled by a constant factor

Table 7.3 provides the calculated production rates of each component within the six operating regions.

The glucose uptake rate can not be determined accurately during the batch phase because the initial substrate is not completely hydrolysed and the measured glucose concentration at 22h is not representative for the actual amount of substrate available at this point. The value for glucose uptake rate given for the two first periods up to 34h is an average value for the entire batch phase.

The uptake of ammonia is very high between 22-34h which is most likely connected to the rapid growth of biomass during this period. Between 34-38h it is relatively low which can be explained by unfavourable growth conditions during this period

begause glucose is exhausted and malate is metabolised. In the last period between 200-203h the ammonia uptake virtually stops reflecting that the biomass growth has stopped at this point. This is again due to substrate starvation.

The enzyme production rate is highest in the period between 38-140h as expected from the substrate limitation in this period. The negative enzyme production rate in the last period is most likely due to an inacurate measurement. It is observed that a considerable enzyme production rate exists in the batch period (0-34h) thus the assumption made in section 5.2 of negligible enzyme production in the batch phase does not hold.

A large production rate of malate is observed during the first period covering up to 22h and a very large consumption rate is observed in the period between 34-38h. A large mannitol formation rate is only seen in the oxygen limited period between 140-200h. The mannitol production rate is practically never negative which implies that mannitol is not significantly metabolised.

The specific uptake and production rates are estimated for each operating period based on the average amount of biomass within the period, the results are presented in table 7.4

Table 7.4. Estimated specific prodcuction rates of each identified component within the six operating regions. The units are 10^{-3} C-moles/g DW/h for compounds containing carbon and 10^{-3} moles/g DW/h for compounds not containing carbon

Period (h)	glc uptake	NH_3	O_2	CO_2	enzyme**	malate	mannitol
0-22	6.51*	1.06	conf.	conf.	0.27	1.16	0.135
22-34	6.51*	1.10	conf.	conf.	0.19	0.047	-0.025
34-38	1.69	0.16	conf.	conf.	0.24	-1.10	0.038
38-140	2.69	0.26	conf.	conf.	0.16	-0.009	0.016
140-200	1.64	0.17	conf.	conf.	0.05	0.023	0.084
200-203	0.11	0.016	conf.	conf.	-0.003	0.022	-0.001

*The glucose uptake rate in these two periods is an average for the time period 0-34h as reliable measurements of substrate concentration could not be obtained during this period

**Due to confidentiality the enzyme yields are scaled by a constant factor

7.2.3 Estimation of yields coefficients

The yield coefficients between the product and growth of biomass can be determined for each operating region by applying a simple mass balance to each operating region. The yield coefficients used in the following are calculated as:

$$Y_{xi}^{k} = \frac{n_i^{k+1} - n_i^k}{n_x^{k+1} - n_x^k} \tag{7.3}$$

where k denotes the current operating region and i denotes the component. The term $n_i^{k+1} - n_i^k$ is the produced number of moles of component i within operating region k.

Table 7.5. Estimated yield coefficients of each identified component with respect to biomass within the six operating regions. All units are moles$_i$/moles$_x$

	glc uptake	NH_3	O_2	CO_2	enzyme**	malate	mannitol
Period (h)	Y_{xglc}	Y_{xn}	Y_{xo}	Y_{xc}	Y_{xp}	Y_{xmal}	Y_{xman}
0-22	2.977*	0.315	conf.	conf.	0.080	0.344	0.040
22-34	2.977*	0.592	conf.	conf.	0.103	0.025	-0.013
34-38	0.879	0.082	conf.	conf.	0.124	-0.571	0.020
38-140	11.10	1.088	conf.	conf.	0.649	-0.036	0.066
140-200	9.177	0.937	conf.	conf.	0.280	0.130	0.471
200-203	0.718	0.105	conf.	conf.	-0.021	0.141	-0.007

*The glucose uptake rate in these two periods is an average for the time period 0-34h as reliable measurements of substrate concentration could not be obtained during this period
**Due to confidentiality the enzyme yields are scaled by a constant factor

Table 7.5 provides the calculated yield coefficients of each component with respect to biomass within the operating regions. Due to the uncertainty on the estimation of glucose utilised at t=22h the yield coefficients for the two first periods for Y_{xglc} are average values for the period between 0 and 34h.

Y_{xglc} reflects how many moles of substrate are taken up per biomass produced. Values below one, as observed between 34-38h and between 200-203h, reflect that more biomass is produced than glucose taken up which is due to metabolism of malate. In the period 200-203h Y_{xmal} indicates that malate is formed, however this value is very sensitive to measurement error. During the most productive periods (38-200h) relatively little biomass and relatively much enzyme is formed per mole of substrate taken up.

7.3 Model for entire cultivation

Based on the results obtained in the previous section combined with the observations made in chapter 6 a model describing the entire cultivation is proposed. Based on the observations it is assumed that a distinct overall metabolism with constant stoichiometric coefficients is valid in each operating region. Only the compounds actually identified in the experiment described in chapter 6 are assumed to be present. The overall stoichiometric equation is written as:

$$X^j + Y^j_{xc}CO_2 + Y^j_{xp}P + Y^j_{xmal}mal + Y^j_{xman}man - Y^j_{xs}S - Y^j_{xn}NH_3 - Y^j_{xo}O_2 = 0 \quad (7.4)$$

where j indicates the model used which is dependent on the operating region. The yield coefficients, $Y^j_{x,i}$, are expressed as the rate of formation of compound i divided by the rate of formation of biomass. In this formulation the carbon balance must obey:

$$Y^j_{xs} = 1 + Y^j_{xc} + Y^j_{xp} + Y^j_{xmal} + Y^j_{xman} \quad (7.5)$$

The reaction rates for this model are given below. All production rates are expressed as C-moles/h or moles/h.

$$\frac{dx}{dt} = \mu^j x \tag{7.6}$$

$$\frac{ds}{dt} = -Y^j_{xs}\mu^j x + F_{dosing} \tag{7.7}$$

$$\frac{dNH_3}{dt} = -Y^j_{xn}\mu^j x + F_{NH_3} \tag{7.8}$$

$$\frac{dp}{dt} = Y^j_{xp}\mu^j x \tag{7.9}$$

$$\frac{dmal}{dt} = Y^j_{xmal}\mu^j x \tag{7.10}$$

$$\frac{dman}{dt} = Y^j_{xman}\mu^j x \tag{7.11}$$

$$CER = Y^j_{xc}\mu^j x \tag{7.12}$$

$$OUR = Y^j_{xo}\mu^j x \tag{7.13}$$

F_{dosing} is the molar carbon feed dosing flow rate given as C-moles/h. F_{NH_3} is the molar feed flow rate of ammonia given as moles/h.

Each model is named A through F corresponding to the operating regions. The characteristics of each model is given below.

Model A This is the first part of the batch phase. This model is valid until a maximum in the filtered OUR/CER signal occurs. At this point it is switched to model B.

Model B This model is valid for the last part of the batch phase. A switch to model C is made when the glucose concentration reaches zero.

Model C This model is valid until the malate concentration reaches zero. At that point a swith to model D is made.

Model D The production phase is contained in this model and it is valid as long the DOT is above 6% and the substrate concentration is above zero. If the DOT drops below 6% a switch to model E is made. If the substrate concentration drops to zero a switch to model F is made.

Model E This model reflects the oxygen limited period. If the DOT increases to above 6% a switch is made to model D. If the glucose concentration drops to zero a switch to model F is made.

Model F This model reflects the substrate starvation period. If the glucose concentration rises above zero a switch to model D is made.

Table 7.6. Specific growth rates and yield coefficients in the model covering the entire cultivation

Model	A	B	C	D	E	F
$\mu^j(h^{-1})$	0.14	0.09	0.05	0.0075	0.005	0.002
Y^j_{xs}	2.977	2.977	0.879	11.10	9.177	0.718
Y^j_{xp}	0.080	0.103	0.124	0.649	0.280	-0.021
Y^j_{xmal}	0.344	0.025	-0.571	-0.036	0.130	0.141
Y^j_{xman}	0.040	-0.013	0.020	0.066	0.471	-0.007
Y^j_{xc}	conf.	conf.	conf.	conf.	conf.	conf.
Y^j_{xo}	conf.	conf.	conf.	conf.	conf.	conf.
Y^j_{xn}	0.315	0.592	0.082	1.088	0.937	0.105

The specific growth rates are assumed to be constant for each model. The parameters of the six models are given in table 7.6.

The definition of the parameters is given in table 7.7.

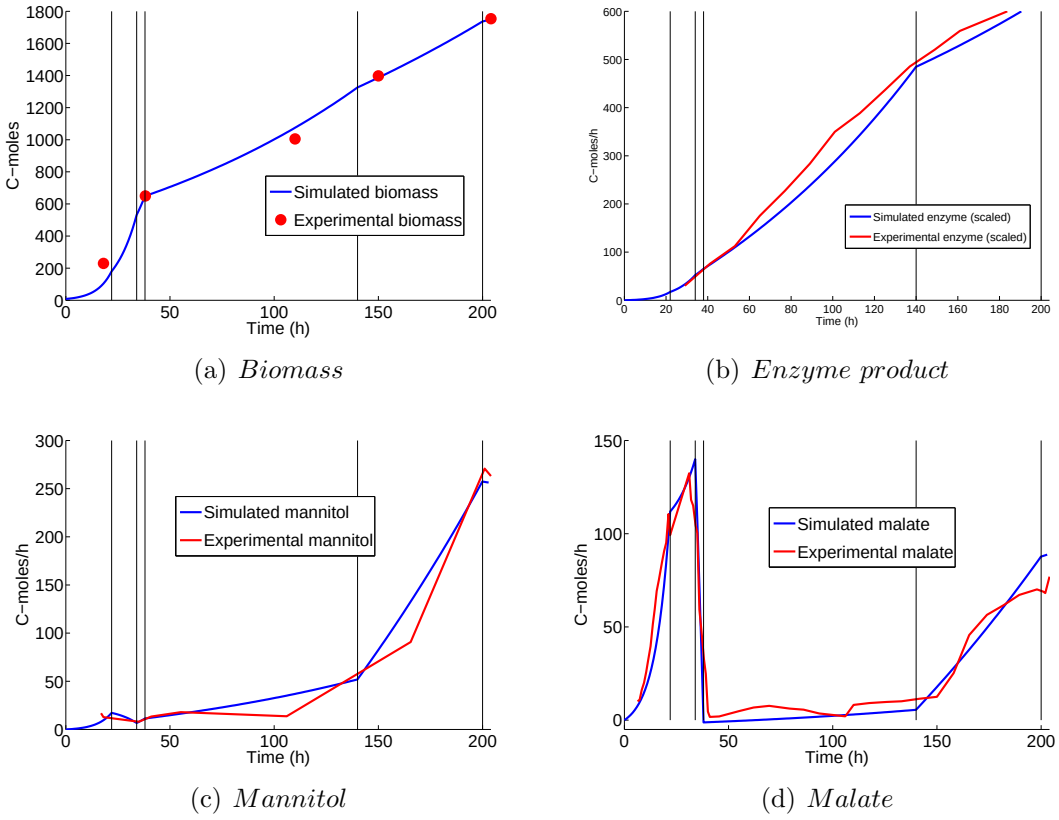
Table 7.7. Definition of parameters in the model for the entire cultivation where j denotes model A through F

Variable	Description	Unit
μ^j	Specific growth rate	$1/h$
Y^j_{xs}	Yield coefficient of biomass to substrate	C-moles s/moles DW
Y^j_{xp}	Yield coefficient of biomass to product	C-moles p/moles DW
Y^j_{xmal}	Yield coefficient of biomass to malate	C-moles mal/moles DW
Y^j_{xman}	Yield coefficient of biomass to mannitol	C-moles man/moles DW
Y^j_{xc}	Yield coefficient of biomass to carbon dioxide	moles CO_2/moles DW
Y^j_{xo}	Yield coefficient of biomass to oxygen	moles O_2/moles DW
Y^j_{xn}	Yield coefficient of biomass to ammnonia	moles NH_3/moles DW
F_{dosing}	Density of dosing feed	C-moles/h
F_{NH_3}	Molar feed flow rate of ammonia	moles/h

7.3.1 Simulation of model for entire cultivation

The proposed model is simulated using the parameters given in table 7.6 combined with the conditions (dosing feed flow etc.) which have atually been applied to the AFF1108 cultivation. Some of the results are shown in figure 7.1 and 7.2.

The horisontal lines indicate the switch between different operating regions. The biomass production and formation of enzyme, malate and mannitol are described very well by the model. However large discrepancies are observed for glucose, CER and OUR. These discrepancies are due to the model structure which implies that formation of products is proportional to the formation of biomass. In the case of glucose (figure 7.2 (c)) this means that the uptake rate increases with time within



(a) *Biomass* (b) *Enzyme product*

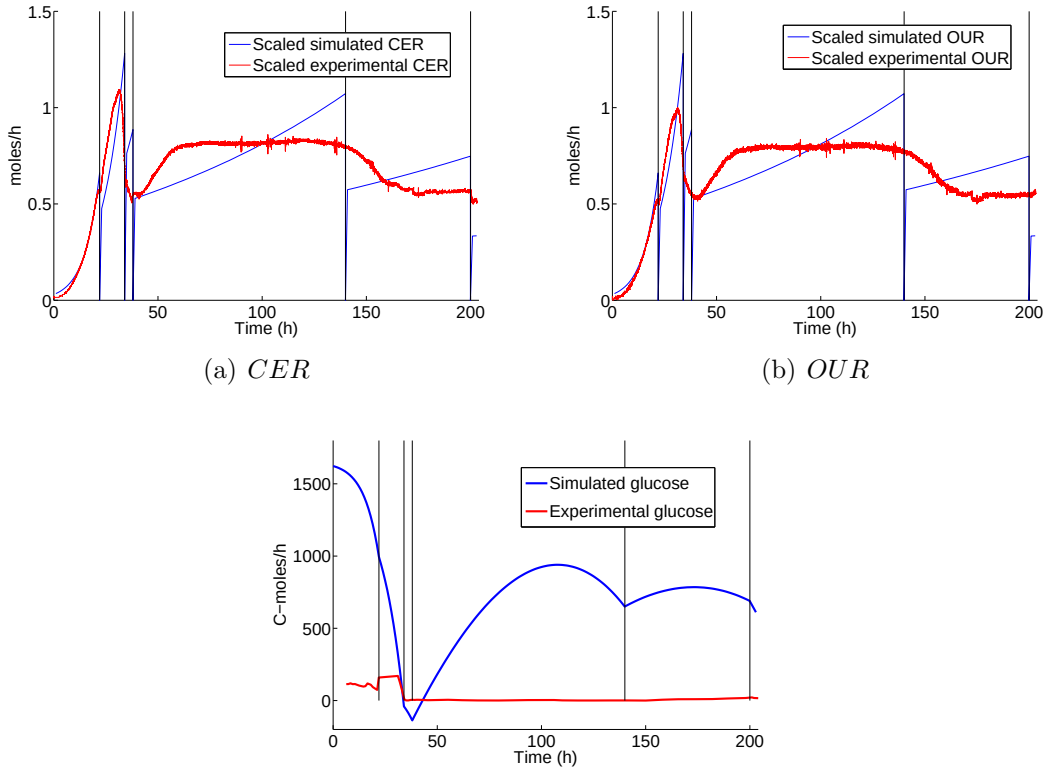
(c) *Mannitol* (d) *Malate*

Figure 7.1. Simulation results performed using the proposed model compared to measured values of the AFF1108 cultivation. The horisontal lines indicate the switch between different operating regions

each operating region due to biomass formation. The dosing feed flow rate however is constant in eg. the period between 38-140h where the glucose concentration is virtually zero indicating a discrepancy between the proposed model structure and the experimental observation. During most of this period CER and OUR (figure 7.2 (a and b)) are constant which is also in conflict with the present model structure.

In order to describe these phenomena with the current model either the yield coefficients have to change significantly within each operating period or the specific growth rate has to decrease significantly. Both of these explanations do not seem reasonable because the growth conditions are in fact rather constant within each of the operating regions.

A more plausible explanation is the occurence of a significant amount of inactive biomass. In literature it is often observed that *Aspergillus oryzae* has large regions with dead hyphae which have no metabolic activity but still contribute in the biomass measurement. A model structure which handles this phenomenon is presented in the next section.



(a) *CER* (b) *OUR*

(c) *Glucose*

Figure 7.2. Simulation results performed using the proposed model compared to measured values of the AFF1108 cultivation. The horisontal lines indicate the switch between different operating regions

7.4 Introducing inactive region of biomass

Based on the model simulations made in section 7.3 a segregated model is proposed which operates with two types of biomass. The first type is active biomass which contributes to the metabolism and the second is inactive biomass which does not contribute to the metabolism. The inactive biomass has originally been active but has become inactive due to severe limitations or simply due to age. In the following it is assumed that the active biomass is converted to inactive biomass by a first order process:

$$\frac{dx_h}{dt} = k^j x \tag{7.14}$$

where k is the rate constant for the first order degradation process, x_h is the number of C-moles of inactive biomass and j indicates the model used. The expression for growth as given in equation 7.6 is now changed to:

$$\frac{dx}{dt} = \mu^j x - k^j x \tag{7.15}$$

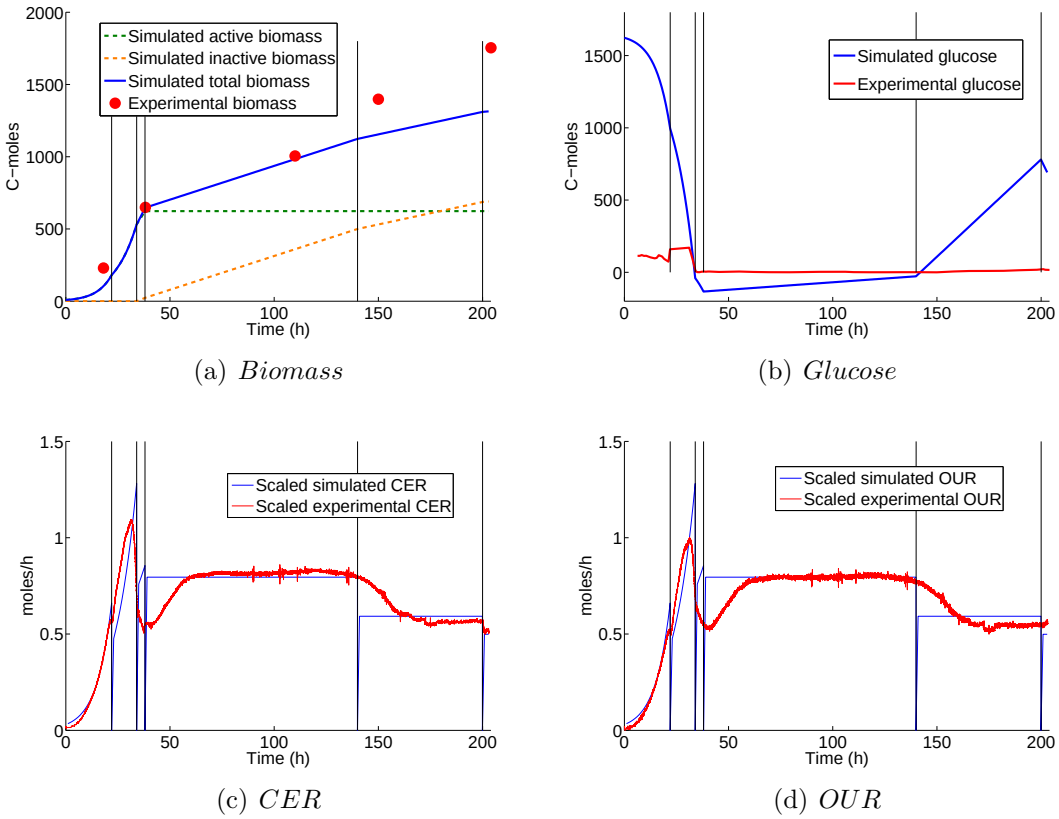
The rest of the model is identical to the one given in section 7.3. All parameters are kept unchanged and the new degradation constants, k, have been fitted by manual

iterations to make the best fit to the experimental results . The values are given in table 7.8 together with μ.

Table 7.8. Values for k and μ for each operating period used in the simulation

Period (h)	k^j	μ^j
0-22	0	0.14
22-34	0	0.09
34-38	0.01	0.05
38-140	0.0075	0.0075
140-200	0.005	0.005
200-203	0.002	0.002

Simulation results using this model are presented in figure 7.3



(a) *Biomass* (b) *Glucose*

(c) *CER* (d) *OUR*

Figure 7.3. Simulation results performed using the modified model including an inactive region of biomass compared to measured values of the AFF1108 cultivation. The horisontal lines indicate the switch between different operating regions

Figure 7.3 shows that the model is still too simple to represent the data correctly but in general a much better performance is achieved. Figure 7.3(a) shows the simulated regions of active and inactive biomass as well as the total amounts of simulated and measured biomass. The specific growth rates and the degradation

constant have only been fitted roughly and the yield coefficients are kept from the original model so there are wide opportunities to improve the performance of this model structure. It is noted that the model is very sensitive to both μ and k. It has been chosen to let the inactivation of biomass start after the batch phase has ended. The simulation indicates that a significant amount of inactive biomass is formed during the cultivation. At the end of the cultivation more than half of the total biomass is inactive.

The substrate concentration is not simulated properly as periods occur where it becomes negative. The glucose concentration is increasing after 140h but the simulated increase is much too steep. The glucose concentration is also very sensitive to the growth rate as a balance between supplied and metabolised matter exists. The reason for the high glucose concentration at the beginning of the batch compared to the measured one is that all substrate is regarded as glucose. In reality only a part of the substrate has converted into glucose at this point.

The simulated OUR and CER now reflect the measured values much better than before. Due to the chosen values of the degradation constant the variables evolve linearly with time during the two periods between 38-140h and 140-200h.

A general observation is that the transitions between operating regions are not reproduced satisfactorily. In reality the transitions are not as abrupt as in the simulation as the conditions change gradually and different regions of the cultivation tank have different concentrations of substrates. A smooth transition between the models is needed to obtain a better performance. However to describe such periods a more detailed model is necessary to handle the switching between pathways in the microorganisms.

7.5 Conclusion

This section has presented a study of the reaction rates of the compounds identified in the AFF1108 batch. A model has been proposed to describe the behaviour of the entire cultivation process. The model has some serious shortcomings as only one type of biomass is assumed to exist. Therefore a more refined model containing both active and inactive biomass has been proposed. This model gives a better description of the observed experimental results but still lacks to handle the transition between operating regions.

A more detailed model will certainly give a better description of the observed experimental results. It must be remembered though that as the model complexity and number of parameters grows the identification of the model is drastically complicated. This is especially the case when a limited number of data sets is available and if the data quality is questioned. The model complexity is also determined by the intended purpose of the model. The models proposed in this chapter can serve as monitoring tools and function as software sensors for estimating the outcome of the cultivation. The current accuracy is however not adequate to serve as process models for tuning advanced control systems.

The observations strongly support the hypothesis that distinct operating regions are present during the cultivation and that a significant amount of inactive biomass is present.

Part IV

Cultivation control perspectives

8

Black-box modelling of production data using a Grid of Linear Models

8.1 Introduction

This chapter presents an application of a black-box methodology for modelling batch process which has been developed in Bonné and Jørgensen (2003). This method is called *Grid of Linear Models* (GoLM) and it is developed for estimation of models for batch and other periodic or continuous operations. The resulting models are large Linear Time Invariant (LTI) models which capture the time varying dynamics of fed-batch processes rather well. In this chapter the GoLM method is used to generate a model for the fed-batch cultivation of *Aspergillus oryzae* studied in this thesis. The model is well suited for implementation in a Model Predictive Control (MPC) framework (Bonné (2005)).

The data used in this chapter are industrial data from one of Novozymes' production facilities. Due to confidentiality the enzyme production levels are scaled by an arbitrary factor.

This chapter is structured as follows: Section 8.2 provides an overview of the variables available from the industrial production plant. In section 8.3 the method for selecting variables to be included in the GoLM model is presented. Section 8.4 presents the model estimation and section 8.5 presents a validation of the obtained model. Finally a conclusion is given in section 8.6.

8.2 Variables available in industrial data set

Production data from an industrial cultivation plant has been made available by Novozymes A/S. The data set contains batches which have been run with the same strain under similar conditions. The basic recipe used is the same for all batches but in some cases operator intervention has been necessary because certain process variables have exceeded the limits specified in the recipe. The data set contains data from a total of 54 batches leading to both satisfactory and unsatisfactory final enzyme activity. 17 of the batches supplied do not contain all the desired variable trajectories and are discarded. Thus only 37 batches have undergone further treatment. By closer inspection 13 of the 37 remaining batches have had problems with one or more of the measured variables. Among these erroneous measurements are: DOT values, gas analysis measurements (CER and OUR) and unreliable enzyme activity measurements. These 13 batches have been discarded to avoid modelling of corrupted data, thus only 24 batches are used for actual modelling. The variables

in these data sets are given in table 8.1.

Table 8.1. Variables available from industrial data set supplied by Novozymes A/S

Variable	Type	Sampling rate
Time	On-line	$6h^{-1}$
Air flow	On-line	$6h^{-1}$
NH_3 flow	On-line	$6h^{-1}$
Accumulated NH_3 flow	On-line	$6h^{-1}$
Dissolved O_2 tension	On-line	$6h^{-1}$
pH	On-line	$6h^{-1}$
Feed flow measured	On-line	$6h^{-1}$
Accumulated feed flow	On-line	$6h^{-1}$
Feed flow set point	On-line	$6h^{-1}$
Back pressure	On-line	$6h^{-1}$
Bottom pressure	On-line	$6h^{-1}$
O_2 uptake rate (OUR)	On-line	$6h^{-1}$
Accumulated O_2 uptake	On-line	$6h^{-1}$
CO_2 evolution rate (CER)	On-line	$6h^{-1}$
Accumulated CO_2 evolution	On-line	$6h^{-1}$
Respirative quotient (RQ)	On-line	$6h^{-1}$
Weight	On-line	$6h^{-1}$
Laboratory pH	Off-line	$2day^{-1}$
Refractive index	Off-line	$2day^{-1}$
Volume percent of mycelia	Off-line	$2day^{-1}$
Enzyme activity	Off-line	$1day^{-1}$

For each batch there is a total of 21 variables, where of 17 are available as on-line measurements whereas 4 are only available as off-line measurements. Three of these are sampled every 12 hours but the enzyme activity, which is the main quality variable for the process, is only measured once every 24 hours. The on-line variables are sampled every 10 minutes and they all contain 1153 samples, corresponding to a batch length of 192 hours. The Oxygen Uptake Rate (OUR) and Carbon dioxide Evolution Rate (CER) are calculated by the control system installed at the cultivation plant. The system also calculates the Respirative Quotient (RQ) as the ratio between CER and OUR.

8.3 Selection of variables for modelling

The supplied data set contains a total of 21 variables (including time) but not all of them contain information which is highly valuable for model estimation, eg. some of the variables may be highly correlated. Inclusion of all 21 variables in the GoLM modelling framework is not feasible, therefore the variables containing the highest information content are to be selected. To facilitate the variable selection Multi-way Principal Component Analysis (MPCA) has been applied Kosanovich *et al.*

(1994); Gregersen (1999). 9 of the 24 batches have been selected for the analysis. All available variables except time have been used in the analysis. In order not to study the behaviour of a standard batch the selected batches have had no operator interference. The feed flow rate has been maintained at the predefined level and the batches have been run according to the specified recipe. The purpose is to gain a further understanding of which process variables are connected to the enzyme formation and responsible for the observed variations in product quality.

As the batches are carried out using a predetermined batch operations model, it is expected that the trajectories of the measurements are very much alike for all batches. Therefore it has been decided that the first step in the analysis is to calculate mean trajectories for all the variables in the 9 batches and subtracting this mean trajectory from the actual trajectory. These deviation trajectories thus represent the difference from the mean batch, see figure 8.1.

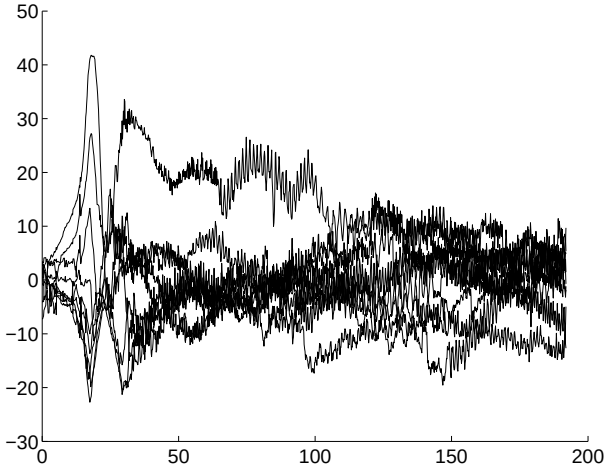


Figure 8.1. Difference trajectories for the 9 batches used for the MPCA analysis. Here DOT is shown, the same procedure has been used for all the variables.

These deviation trajectories have been stored in a three-way matrix X ($J \times K \times I$), where J is the number of variables, K is the number of samples from each batch, and I is the number of batches. In this case the dimensions are $20 \times 1153 \times 9$, as time is not included in the analysis. The matrix X is unfolded to a two-way matrix as indicated in figure 8.2. This two-way matrix is called X ($J \times IK$). Each column in X represents a certain variable for all batches and all points in time. Each column is mean centred and scaled to unit variance.

The variance captured by each principal component (PC) is shown in figure 8.3. The two first PC's capture a relatively large part of the variance and the variance captured by PC 3 and 4 is significantly lower. Another drop appears from the 4th to the 5th PC and from here the decrease is almost linear. The cumulated variance captured by the 4 first PC's is 54.8 %.

Figure 8.4 shows the loadings of each variable on the two first principal components. It is seen that the enzyme activity has a relatively small loading on both PC 1 and PC 2, indicating that it is only responsible for a small explanation of the entire variation of the data. The variables having the highest loadings on PC

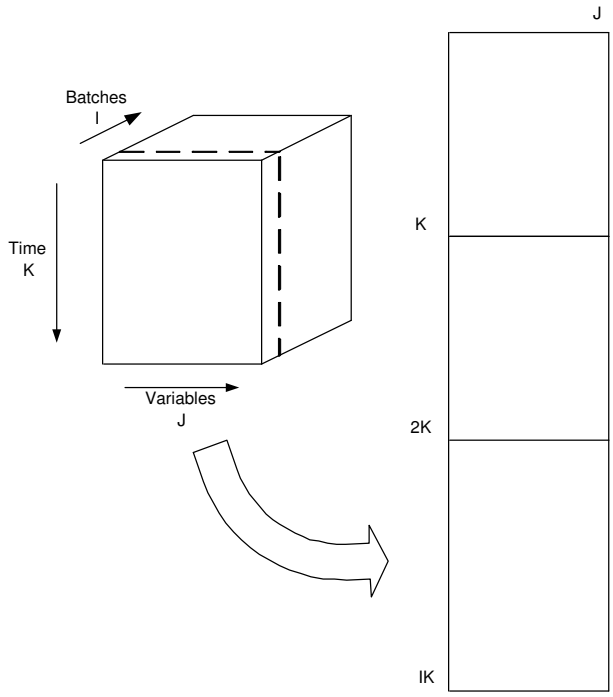


Figure 8.2. The difference trajectories have been organised into a 2-dimensional structure and are unfolded by stacking them underneath each other.

1 are accumulated ammonia flow (negative loading), weight, accumulated feed flow and accumulated CER. This shows that a large part of the deviation from the mean batch originates from accumulated variables. This is not surprising as a higher total substrate feed results in a higher weight of the cultivation broth, a higher conversion of substrate and consequently a higher oxygen demand. The lack of correlation between the accumulated variables and the enzyme activity indicates that the product yield, ie. the ratio between enzyme produced and substrate consumed, is not constant.

An interesting observation is that there is a negative correlation between enzyme activity and DOT. This suggests that it is beneficial for the product formation to follow a relatively low DOT trajectory. This hypothesis is supported by investigating PC 3 and 4 (figure 8.5). Here it is seen that the enzyme activity has a negative loading on PC 3 and that DOT has a positive loading on the same PC, they are thus negatively correlated.

The selection of variables for modelling is based on investigation of the loadings of each variable on the first four PC's combined with process knowledge. The goal has been to keep the number of variables as low as possible but still obtain good predictive capabilities and capture as much of the variation from the mean batch as possible. To keep the model as simple as possible accumulated gas analysis measurements have not been included. The actual CER and OUR measurements are used instead in order to capture some dynamic phenomena in the system, eg. conversion of a suddenly changing feed flow rate. The only input used in the model is the feed flow set point which is because it is the controlled variable having the largest impact on the process and is intended to be used as the actuator (manipulated

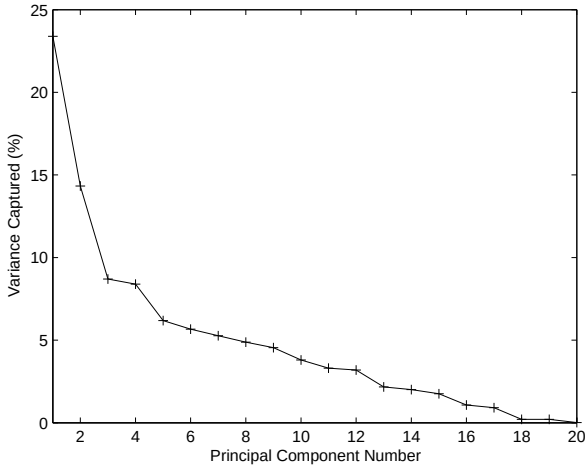


Figure 8.3. Variance captured by each principal component.

variable) for MPC control. The variables chosen for the model are given in table 8.2.

Table 8.2. Variables used for modelling including the cut-off frequencies used in the fourth order Butterworth filter

Variable	Type	Cut-off frequency/Nyquist frequency)
Feed rate	Input	0
DOT	Output	0.1
EnzA	Output	0.1
Weight	Output	0.05
OUR	Output	0.1
CER	Output	0.1
Ammonia flow	Output	0.8
Air flow	Output	0.8

8.4 Modelling

The raw batch data has been filtered using a Butterworth filter of 4th order with cut-off frequencies as shown in table 8.2. The last few hours of the data set have been truncated because the emptying of the cultivation tank has started at this time for some of the batches. It is not desired to model these phenomena because the actual cultivation process has terminated at this point. The data has been truncated to a duration of 190 hours. The cut-off frequencies have been tuned carefully for each variable in order to filter away as much noise as possible and still not lose too much information about the process dynamics. The number of variables has been kept at a minimum in order not to render the model too complicated and only one input (to be used as the manipulated variable in the controller) has been chosen.

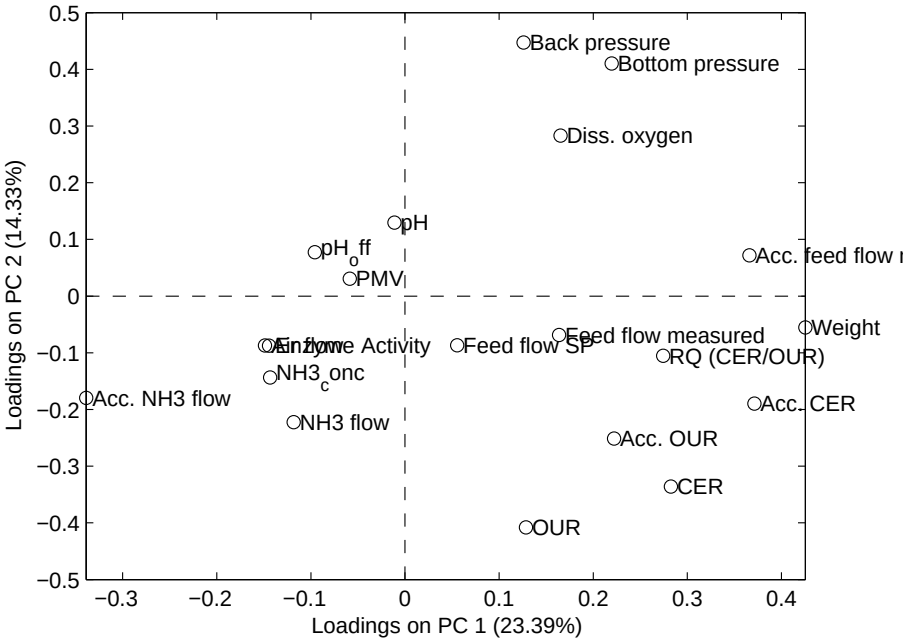


Figure 8.4. Loadings of all variables on the two first principal components.

The 24 batches have been divided into 16 modelling and 8 validation batches. Both modelling and validation batches have had some operator intervention on the feed flow set point. The maximum order for each output/output and input/output relationship has been set to 4.

112 grid points have been used to model the process from t=0 hours to t=190 hours. For the first 60 hours the grid point spacing has been set to 1 hour in order to cover the very dynamic behaviour of the batch phase. For t=62.5 hours to t=190 hours the grid point spacing has been increased to 2.5 hours as the process is less dynamic during the fed-batch phase.

8.5 Model validation

The estimated orders of the individual ARX models representing each relationship between inputs and outputs are given in table 8.3. The model orders do not have a direct physical interpretation but it can be noted that the relation between the feed flow set point (input) and all the outputs have high model orders, supporting the fact that variations in feed flow play a fundamental role for the process.

The resulting GoLM model can be represented as a State Space (SS) model. This representation has been used for the following simulations. State estimation is carried out at each grid point by using the actual measurements obtained at that point in a Kalman filter. An initial estimate of the system state is obtained by estimating the initial state (t=0) using the initial measurements available at this time point. The pure simulation (PS) is obtained by using only this initial estimated state and all future states are calculated based on this. The one step ahead (OSA) simulation

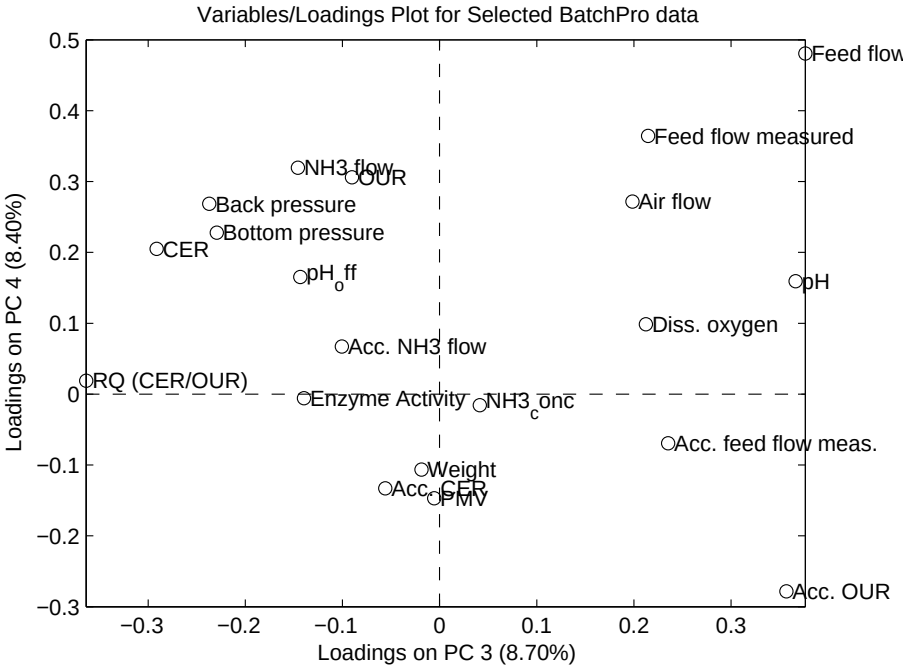


Figure 8.5. Loadings of all variables on the third and fourth principal component.

is obtained by using the state estimated at each grid point as initial value to predict the outputs at the next grid point.

The 8 validation batches have been simulated using the model developed from the 16 modelling batches. Results for one of these validation simulations is given in figure 8.6 to 8.13. Figure 8.6 shows the input signal used for the simulation and the input signal which has been applied to the reference batch. It is seen that the reference input reaches a constant level after around 40 hours. The input for the simulated system (actual input) also reaches this level but is manipulated further by intervention from process operators after around 70 hours.

The enzyme activity (figure 8.7) is simulated rather well by the model. The reference is significantly lower than the actual and it is seen that the model is able to capture the change in process conditions and predict a higher enzyme activity the reference batch. The pure simulation follows the reference trajectory closely up to around 35 hours. It then increases and follows the measured trajectory. The pure simulation of the enzyme activity is especially important for control purposes because no measurements are available during the process to give information on the actual activity.

Figure 8.8 shows the prediction of the dissolved oxygen tension. The actual trajectory deviates somewhat from the reference and the pure simulation predicts something in between for most of the simulation. The one step ahead prediction follows the actual trajectory very well. The pure simulation is less important in this case than for the enzyme activity because the measurement is available on-line but the controller still needs a reliable model for prediction of the future behaviour of the process.

Figure 8.9 and 8.10 show the predicted OUR and CER respectively. Some devia-

Table 8.3. Model orders for the identified ARX models. Combination of outputs and inputs (rows) onto outputs (columns).

	DOT	EnzA	Weight	OUR	CER	NH_3 flow	Air flow
DOT	1	1	2	1	1	0	4
EnzA	2	1	0	0	4	0	0
Weight	0	0	4	4	4	1	1
OUR	0	1	4	1	1	4	0
CER	1	4	4	1	1	2	0
NH_3 flow	1	0	3	4	1	4	4
Air flow	3	0	0	1	0	4	3
Feed flow set point	4	4	4	4	3	2	4

tions occur for the pure simulation from 30 to 80 hours for both variables.

The ammonia flow rate is shown in figure 8.11. Here the model is able to predict the trajectory very well, both for one step ahead and pure simulation.

The prediction of the air flow rate (figure 8.12) is quite poor, mainly because it is caused by disturbances of the pressure in the air supply line which can not be modelled. The reason for modelling this variable is to attempt to incorporate information on the amount of air supplied to the system in the model.

The weight of the cultivation broth (figure 8.13) is predicted very well due to the fact that the feed flow information is available but substrate conversion and water evaporation also play a role.

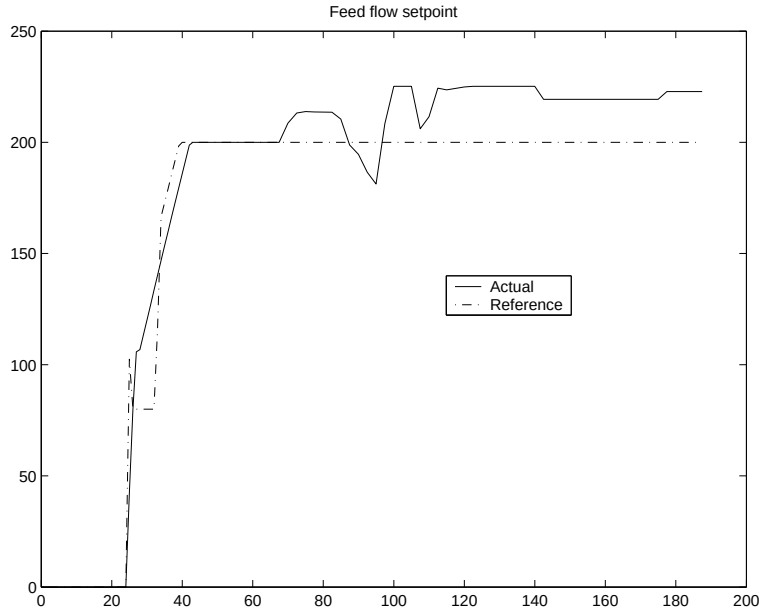


Figure 8.6. Feed flow set point (arbitrary unit). Actual: Simulated trajectory. Reference: Trajectory used in the reference batch.

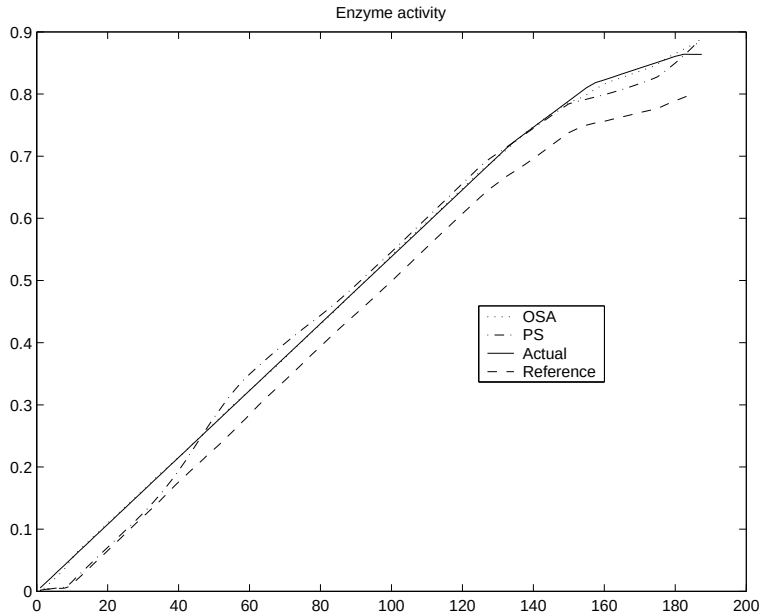


Figure 8.7. Enzyme activity (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

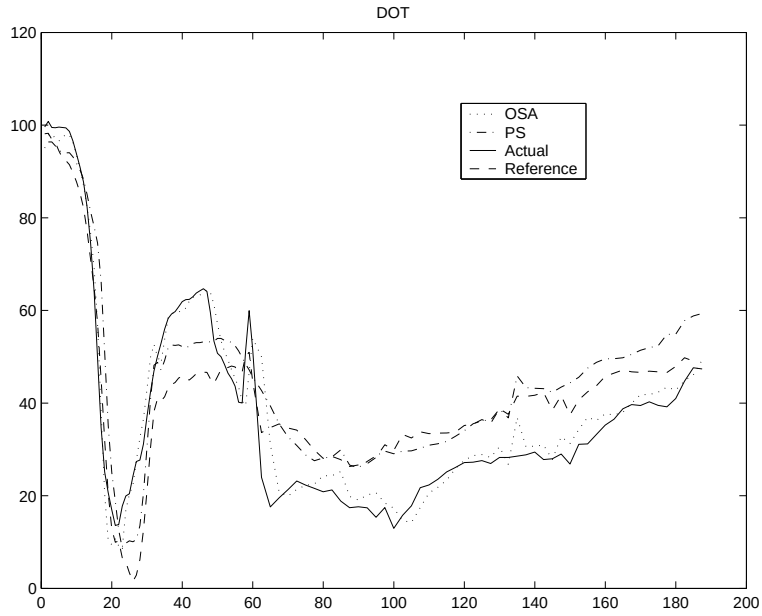


Figure 8.8. Dissolved oxygen tension (per cent of saturation). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

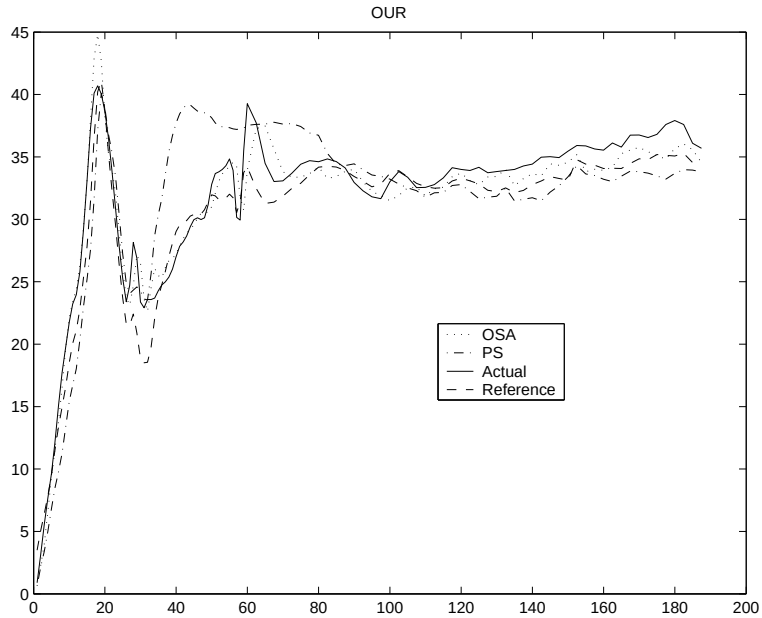


Figure 8.9. Oxygen uptake rate (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

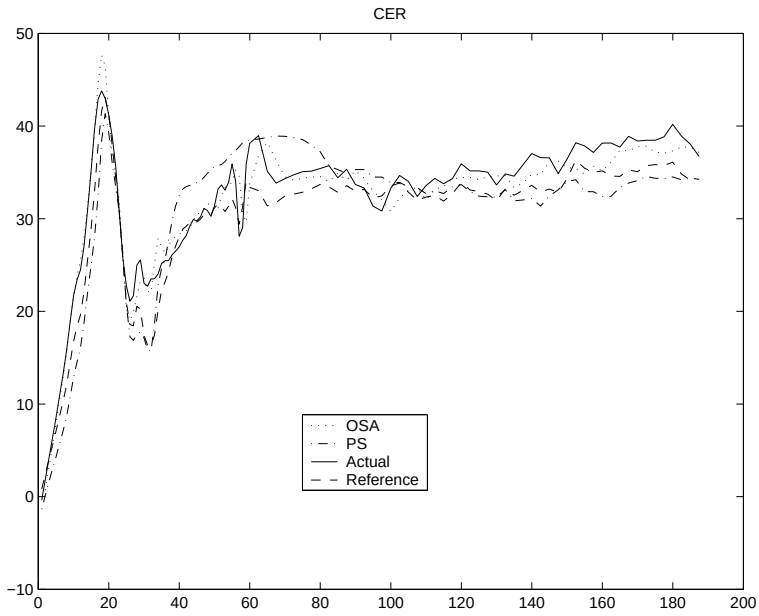


Figure 8.10. Carbon dioxide evolution rate (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

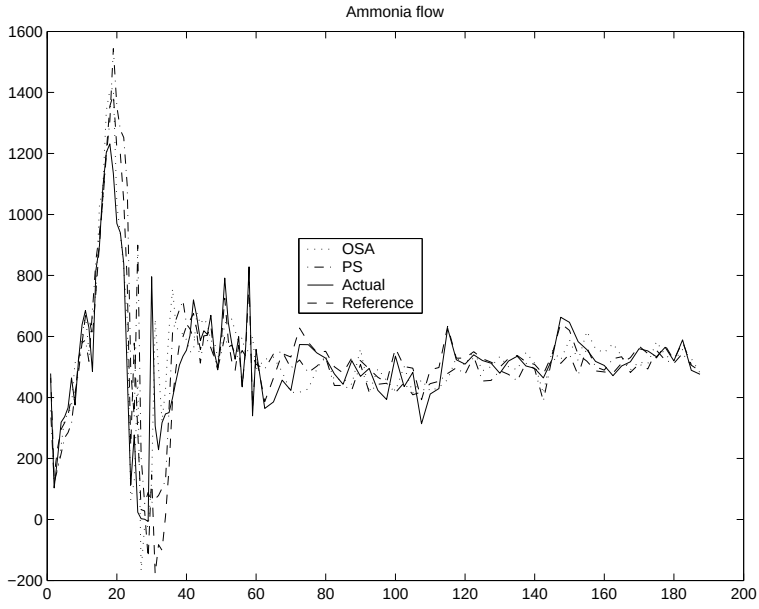


Figure 8.11. Ammonia flow rate (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

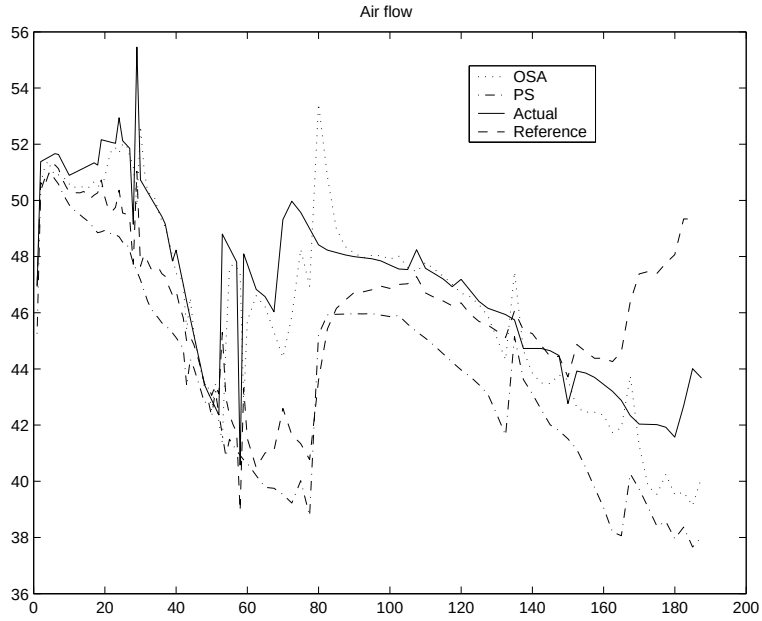


Figure 8.12. Air flow rate (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

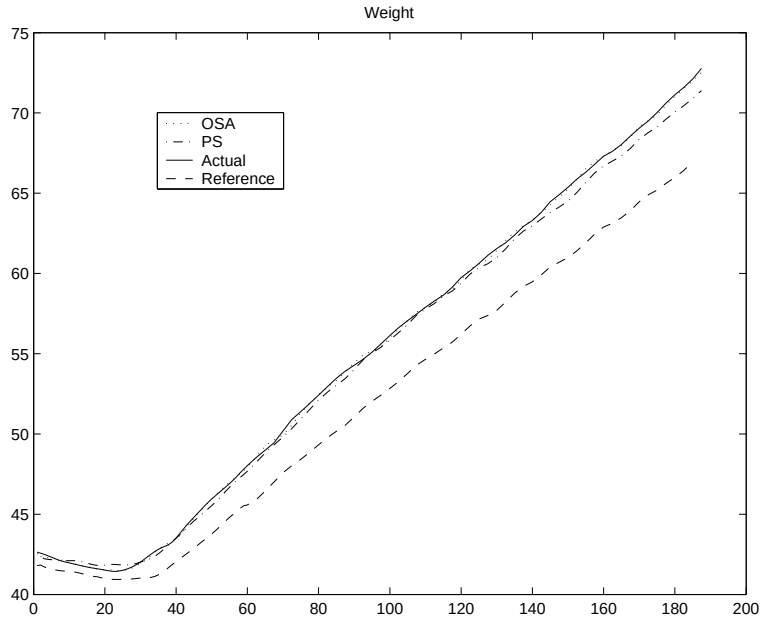


Figure 8.13. Weight (arbitrary unit). OSA: One step ahead prediction. PS: Pure simulation. Actual: Simulated batch. Reference: Reference batch.

8.6 Conclusion

The GoLM modelling framework has been applied to model industrial production data from a fed-batch cultivation process. The modelling framework divides the entire batch duration into a large number of grid points and local linear models are fitted to each of those points. The combination of all these models is able to approximate the behaviour of the entire process. The variable selection has been facilitated by application of multi-way principal component analysis on the data, identifying the parameters having the largest impact on the process dynamics. The developed model generally possesses reasonably good predictive capabilities and the model structure makes it well suited for implementation in a MPC framework.

9

Control of pilot plant cultivation

9.1 Introduction

This chapter provides perspectives on how automatic control can be introduced in industrial fed-batch cultivations. Chemical and biochemical processes will always be sensitive to variations in raw material properties, operating conditions and other disturbances on the process. This sensitivity necessitates some kind of process control in order to achieve a desired reproducible quality of the final product. The control scheme currently used for the cultivation process studied in this work is based on manual intervention by process operators. The inputs to the process are altered if one or more variables are outside some predefined range or if the operators find it necessary for some other reason. Such a scheme depends heavily on the operators attention on the critical process variables as well as their experience. Introduction of a reliable automated control scheme can yield large potential benefits.

The GoLM modelling framework has been succesfully integrated in a Linear Model Predictive Control framework Bonné (2005). The obtained controller has been implemented in the control system at the Novozymes Pilot Plant and a controlled cultivation has been carried out.

This chapter is structured as follows: Section 9.2 gives an overview of the properties and development of the GoLM model used in the controller framework. Section 9.3 introduces the Model Predictive Control framework applied. An experiment has been carried out to evaluate the controller and a description of the controller and the results obtained from this cultivation are given in section 9.4. Section 9.5 concludes this chapter.

9.2 Model development

The GoLM modelling framework (chapter 2) has been used to identify the data-driven model used in the controller framework.

The GoLM modelling framework provides wide possiblities to choose among inputs, outputs, number of grid points etc. A number of different GoLM models have been developed using the eight batches, AFF1082 through AFF1108, available in the pilot plant data set (appendix A provides a thorough description of this dataset). Among the models estimated the most promising has been selected for implementation in the controller framework. Information on the other GoLM models which have been developed but are not selected for implementation in the controller is given in Petersen (2006). The model chosen for controller implementation operates with four outputs and one input. The outputs are integrated ammonia feed flow,

DOT, CER and OUR. The input is the set point for the dosing feed flow rate. The rationale for choosing the four outputs is that they are all available as on-line measurements and provide important information regarding the state and evolution of the cultivation. Also this combination of outputs provides reasonable predictive capabilities. The input has been chosen to be the dosing feed flow rate because it is the process variable which has the largest impact on the process.

It has been chosen to align and truncate the data from the eight batches such that only the fed-batch part of the cultivation has been included. The reason for this choice is that experience shows that modelling the transition between batch and fed-batch mode is particularly challenging because the transition occurs at different points in time from batch to batch which means that the same grid point model is not necessarily valid at the same time point for every batch which is an assumption made the GoLM methodology. The GoLM methodology assumes that the process is time varying within the batch but time invariant from batch to batch. Modelling only the fed-batch part circumvents this issue and it is thus assumed that the fed-batch part of the process is time invariant from batch to batch.

The selected model has the ARX (AutoRegressive with eXogenous inputs) model structure and Tikhonov regularisation has been used in the parameter estimation. The sampling interval is 1/2 hour and the number of grid points is 261 which means that the model covers the 130 hours after feed start. The choice of number of grid points is a trade-off between model granularity and computation time for the model estimation. The computation time is also heavily affected by the maximum model order which can be assigned to each output/output and input/output connection and therefore a maximum of two has been used.

The eight batches have had different initial amounts of substrate which is accounted for in the model by adding the initial substrate concentration as an Extra Initial Condition (EIC). Included in the GoLM framework is the possibility to state such EIC's which in principle are treated as additional initial inputs to the model (Bonné (2005)).

The model orders determined during model estimation for the input/output connections for this GoLM model are summararised in table 9.1.

Table 9.1. Model orders for the output/output and input/output connections in the GoLM model used in the controlled experiment AFF1109

	Int. NH_3	DOT	CER	OUR	Feed flow rate	EIC
Int. NH_3	1	0	2	2	2	2
DOT	1	2	2	0	1	0
CER	2	0	1	0	2	2
OUR	1	1	2	1	1	2

Table 9.1 shows that the feed flow dosing rate is coupled directly to all four outputs which seems reasonable since all outputs are directly affected by changes in the rate of dosing feed flow also the EIC has a direct impact on three of the outputs.

The six batches AFF1082, AFF1098, AFF1099, AFF1100, AFF1101 and AFF1102 have been used for model identification and estimation, batches AFF1103 and

AFF1108 have been used for model validation. Figure 9.1 to 9.4 illustrate the pure simulation prediction of a validation batch (AFF1103) based on the trajectory of one of the model identification batches (AFF1102). Pure simulation implies that only the initial outputs and the actually applied input trajectory is used in the simulation. Under practical control circumstances the online variables will be measured continuously and the estimation will be brought back on track during the cultivation process.

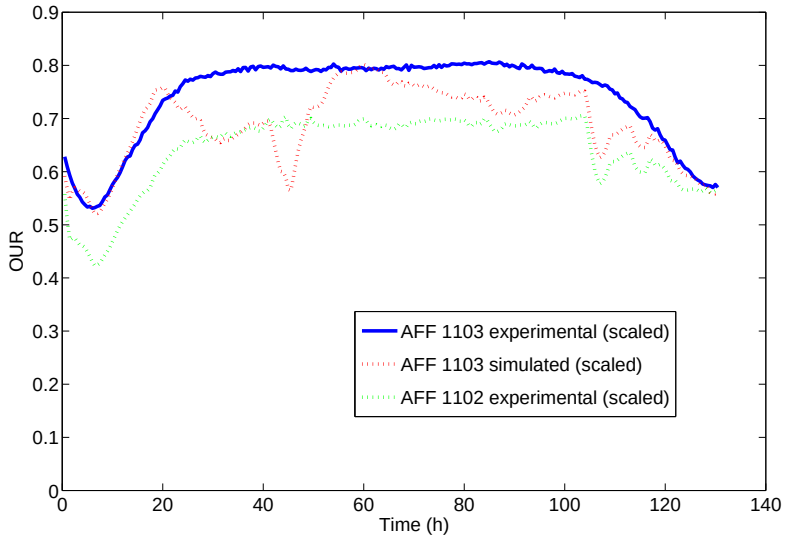


Figure 9.1. Difference trajectories for OUR simulated by the GoLM models used in the controlled experiment AFF1109. AFF1102 is used as the reference batch (experimental trajectory shown) and AFF1103 is the simulated batch (experimental and simulated trajectory shown). All trajectories are scaled by the same factor

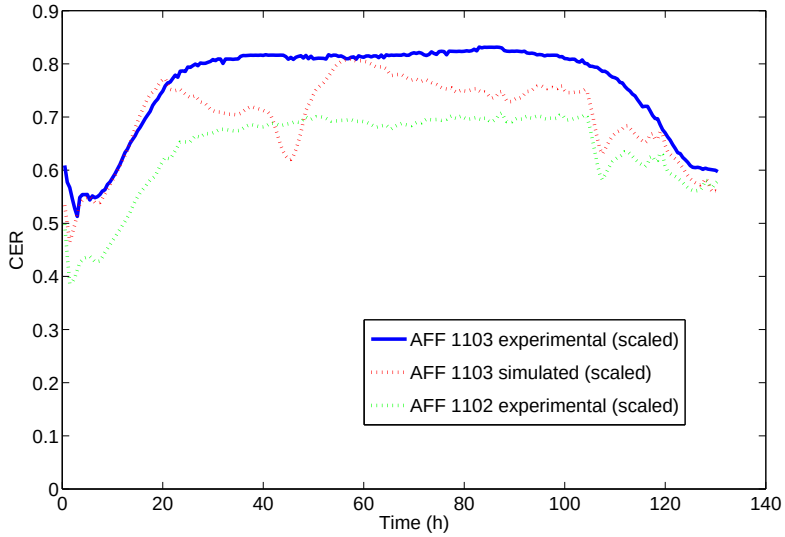


Figure 9.2. Difference trajectories for CER simulated by the GoLM models used in the controlled experiment AFF1109. AFF1102 is used as the reference batch (experimental trajectory shown) and AFF1103 is the simulated batch (experimental and simulated trajectory shown). All trajectories are scaled by the same factor

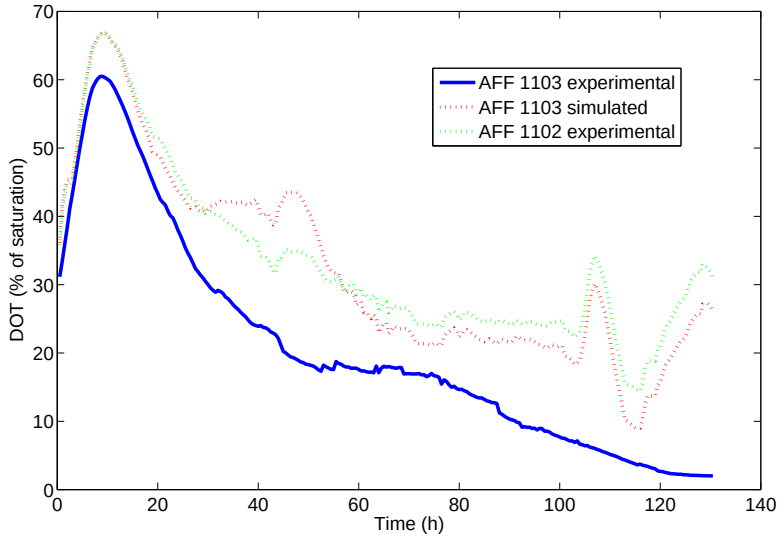


Figure 9.3. Difference trajectories for DOT simulated by the GoLM models used in the controlled experiment AFF1109. AFF1102 is used as the reference batch (experimental trajectory shown) and AFF1103 is the simulated batch (experimental and simulated trajectory shown)

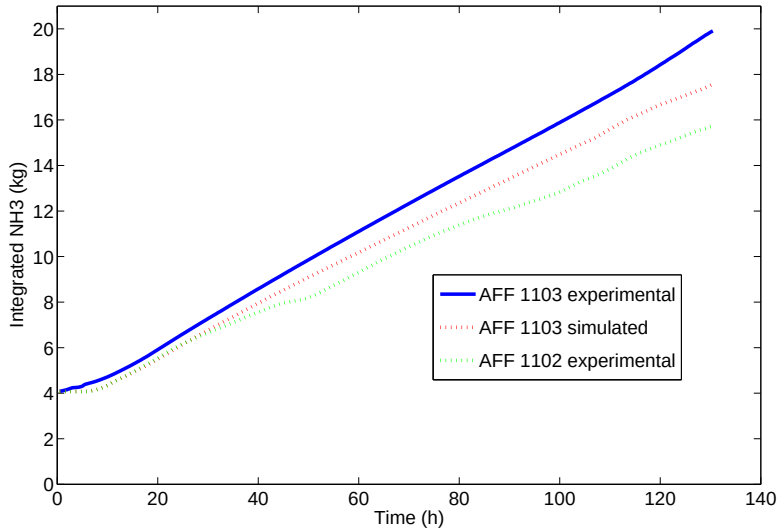


Figure 9.4. Difference trajectories for accumulated ammonia simulated by the GoLM models used in the controlled experiment AFF1109. AFF1102 is used as the reference batch (experimental trajectory shown) and AFF1103 is the simulated batch (experimental and simulated trajectory shown)

Figure 9.1 to 9.4 reveal that the GoLM model generally is able to predict the pure simulation variable trajectories rather well. The predictions for OUR (figure 9.1) and CER (figure 9.2) are rather similar. During the first 20 hours the simulation is very close to the experimental result, during 20 to 60 hours the simulation underestimates the values and after 60 hours there is a tendency to drift towards the reference batch. The DOT level (figure 9.3) is consequently underestimated by the model and especially the last part (from 80 to 130 hours) is not predicted very well by the model. The last simulated variable, the integrated ammonia flow, is shown in figure 9.4. The simulation underestimates the level but the model is able to capture that the level is higher for the simulated batch than for the reference batch. In general there is a tendency for the simulated trajectories to be too close to the ones of the reference batch. Despite some unaccuracies the model is suited for LMPC controller adaptation due to the fact that estimations for the short estimation horisons are rather good and that only online variables are estimated which means that the state estimates are updated continuously.

9.3 Model Predictive Control framework

This section provides a brief introduction to the Model Predictive Controller framework applied in this work.

Model Predictive Control (MPC) is a state of the art controller based on linear state space models and it has already found widespread application in the industry for control of continuously operating processes. MPC has a number of advantages compared to classical control frameworks, multivariable control problems are handled naturally as well as changes in set-points and operation modes. Furthermore both actuator limits and other system constraints are handled. This allows for operation closer to the system constraints which increases the overall profitability of the process (Maciejowski (2002)).

Model Predictive Control is extensively covered in literature, eg. Maciejowski (2002), Muske and Rawlings (1993), Mayne *et al.* (2000), Morari and H. Lee (1999) and Rao and Rawlings (2000).

Figure 9.3 illlustrates the control hierarchy which can be applied to the controller framework described in this work. The upper level in the control hierarchy concerns the overall scheduling of batches and the decisions on this level are ultimately dictated by market demands. Once a certain batch has been planned the execution is handled by the multivariable control level. The MPC controller operates on this level and provides set-point trajectories for underlying SISO (Single Input Single Output) controllers which are connected to the actual process actuators. The MPC controller uses an internal model to recalculate the set-point trajectories with a relatively low frequency (often in the range of hours) whereas the SISO controllers operate with a much higher frequency, often in the range of seconds. The process monitoring level utilises measurements from the process, which are usually identical to the measurements used for feedback in the SISO controllers, and uses this information to estimate the state of the system in order to determine if the process is ready to switch to the next phase of operation. Furthermore the process monitoring level monitors whether the conditions for the multvariable control remain satisfied.

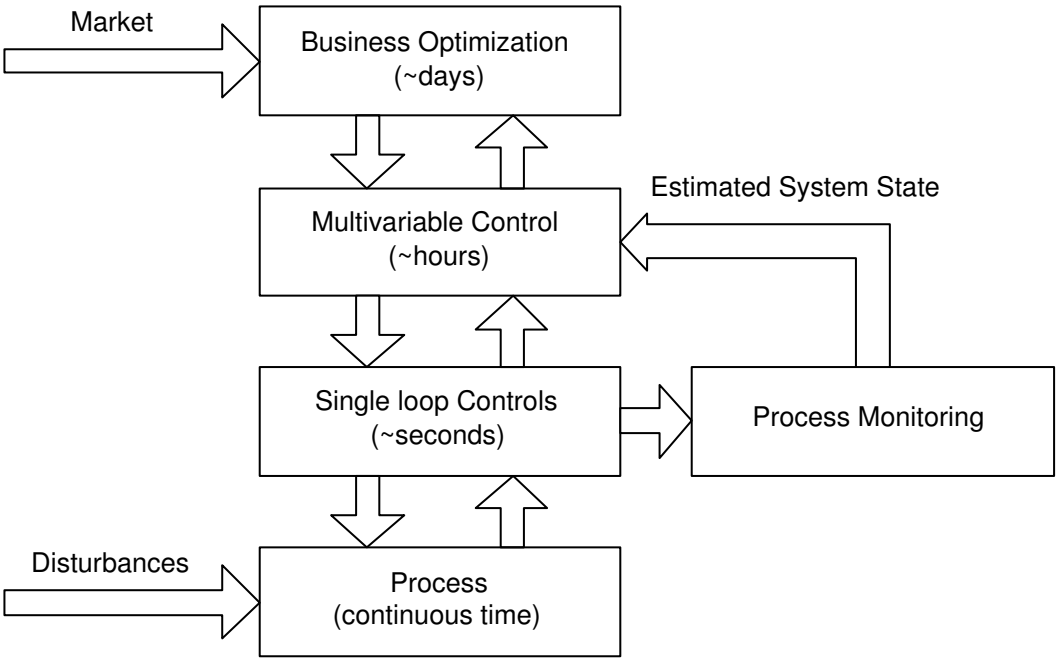


Figure 9.5. The arrangements between multivariable and single loop control in the overall control hierarchy

If that is not the case the multivariable control must be turned off - perhaps only temporarily. The process itself is continuosly subject to disturbances which requires communication in both ways between the control levels in the entire control hierarchy.

9.3.1 Mathematical formulation of MPC

The MPC controller is formulated mathematically to minimise an objective function subject to defined constraints. The standard MPC formulation uses Linear Time Invariant (LTI) models but can be extended to treat Linear Time Varying (LTV) models as well. In this work a development of the MPC controller known as Learning MPC (LMPC) (Bonné (2005)) has been applied. The basic idea in this controller is to combine the asymptotic batch to batch (inter-batch) convergence of Iterative Learning Control (ILC) with the closed loop intra-batch performance of MPC. This way random intra-batch disturbances are handled and the effect of persistent inter-batch disturbances is asymptotically eliminated. In this formulation the system is described as a stochastic LTV system

$$\begin{aligned} x_{k,t} &= A_t x_{k,t-1} + B_t \Delta u_{k,t-1} + E_t v_{k,t} \\ e_{k,t} &= \bar{y}_t - y_{k,t} \\ &= e_{k-1} - C x_{k,t} \end{aligned} \tag{9.1}$$

for $t = 1, ..., N$ with the initial condition

$$\boldsymbol{x}_{k,0} = -C v_{k,0} \tag{9.2}$$

This means that the state $x_{k,t} \in \mathbb{R}^{n_x}$ of system (9.1) at time t in batch k, is given by linear mappings of the state $x_{k,t-1}$ at time $t-1$ in batch k, the control correction $\Delta u_{k,t-1} = u_{k,t-1} - u_{k-1,t-1} \in \mathbb{R}^{n_u(t-1)}$ and a zero-mean Gaussian disturbance $v_{k,t} \in \mathbb{R}^{n_y}$. The tracking error $e_{k,t} \in \mathbb{R}^{n_y}$ of system (9.1) is given as the difference between the system output $y_{k,t} \in \mathbb{R}^{n_y}$ and the desired output reference $\bar{y}_t \in \mathbb{R}^{n_y}$ (Bonné (2005)).

This leads to the following LMPC formulation

$$\begin{aligned}\{\Delta u_{k,l,-1}\}_{l=0}^{N-1} = \arg \min_{\{\Delta u_{k,i}\}_{i=0}^{N-1}} & \left[\sum_{i=1}^{N} \hat{e}'_{k,i|t} Q_{k,i} \hat{e}_{k,i|t} + \Delta u'_{k,i-1} R_{k,i} \Delta u_{k,i-1}\right] \\ s.t. \quad & \hat{x}_{k,i|-1} = A_i \hat{x}_{k,i-1|-1} + B_i \Delta u_{k,i-1} \\ & \hat{e}_{k,i|-1} = \hat{e}_{k-1,i|N} - C\hat{x}_{k,i|-1} \\ & u_{\mathtt{min}\, i-1} \leq \Delta u_{k,i-1} + u_{k-1,i-1} \leq u_{\mathtt{max}\, i-1} \\ & y_{\mathtt{min}\, i} \leq \bar{y}_i - \hat{e}_{k,i|-1} \leq y_{\mathtt{max}\, i}\end{aligned} \tag{9.3}$$

if the weighting matrices Q_k and R_k are block diagonal

$$\begin{aligned}\boldsymbol{Q}_k &= \mathtt{diag}(Q_{k,t}) \\ \boldsymbol{R}_k &= \mathtt{diag}(R_{k,t})\end{aligned} \tag{9.4}$$

and the weighting matrices $\boldsymbol{Q}_{k,t}$ and $\boldsymbol{R}_{k,t}$ are all symmetric and positive definite. Bonné (2005) proves that the optimal solution $\Delta \boldsymbol{u}_{k,-1} = \{\Delta u_{k,l,-1}\}_{l=0}^{N-1}$ to the LMPC problem (9.1) guarantee convergence.

The cost fuction J where $t = -1, 0, \ldots, N$ and $-1 \leq j < N$ is then defined as (Bonné (2005))

$$J_k(\hat{e}_{k,t}(\Delta u_{k,j}), \Delta u_{k,j}) = \|\hat{e}_{k,t}(\Delta u_{k,j})\|^2_{Q_k} + \|\Delta u_{k,j}\|^2_{R_k} \tag{9.5}$$

9.3.2 State estimation

Prior to the evaluation of the objective function it is necessary to perform an estimation of the system states to minimise the influence of measurement noise. Different state estimators exist but most often the Kalman filter is used due to its unique features of providing the statistically optimal estimates (Grewal and Andrews (2001)).

The Kalman filter employed in this work is formulated as in Bonné (2005)

$$\begin{aligned}\hat{x}_{k,t|t} =& \hat{x}_{k,t|t-1} + K_{k,t}(z_{k,t} - C\hat{x}_{k,t|t-1} - \hat{y}_{k-1,t|N}) \\ \hat{x}_{k,t|t-1} =& A_t^S \hat{x}_{k,t-1|t-1} + B_t \Delta u_{k,t-1} + E_t S_t R_{\epsilon_t}^{-1}(z_{k,t-1} - \hat{y}_{k-1,t-1|N})\end{aligned} \tag{9.6}$$

for $t = 0, 1, \ldots, N$, with initial condition

$$\hat{x}_{k,0|-1} = 0 \tag{9.7}$$

The Kalman filter gain matrix $K_{k,t}$ and state estimate covariance matrix $P_{k,t|t}$ are propagated by the following recursions

$$\begin{aligned} P_{k,t|t} &= P_{k,t|t-1}(\boldsymbol{I} - C'K'_{k,t}) \\ K_{k,t} &= P_{k,t|t-1}C'(CP_{k,t|t-1}C' + R_{\epsilon_t} + \bar{P}_{k-1,t|N})^{-1} \\ P_{k,t|t-1} &= A_t^S P_{k,t-1|t-1}A_t^{S'} + E_t\Sigma_t^S E_t' \end{aligned} \tag{9.8}$$

with initial condition

$$P_{k,o|-1} = C'\Sigma_0 C \tag{9.9}$$

In situations where there are no observations of the initial conditions, which is the most common situation for a bio-chemical batch process, the initial state estimate becomes

$$\begin{aligned} \hat{x}_{k,0|0} &= 0 \\ P_{k,0|0} &= C'\Sigma_0 C \end{aligned} \tag{9.10}$$

9.4 Controlled cultivation experiment

The MPC controller described in section 9.3 has been implemented at the Novozymes Fermentation Pilot Plant and it has been used for performing a controlled cultivation. The GoLM model described in section 9.2 has been employed in this control framework.

9.4.1 Controller implementation

The MPC framework has been implemented in Matlab files on a computer in the operating control room at the Novozymes Fermentation Pilot Plant. This computer is connected directly to the DeltaV control system which allows for immediate execution of the control signals provided by the MPC controller.

The controller parameters used in the controlled cultivation are given below. The parameters for the Q matrix are given in table (9.2)

Table 9.2. Output weighting matrix, Q, for the controlled experiment

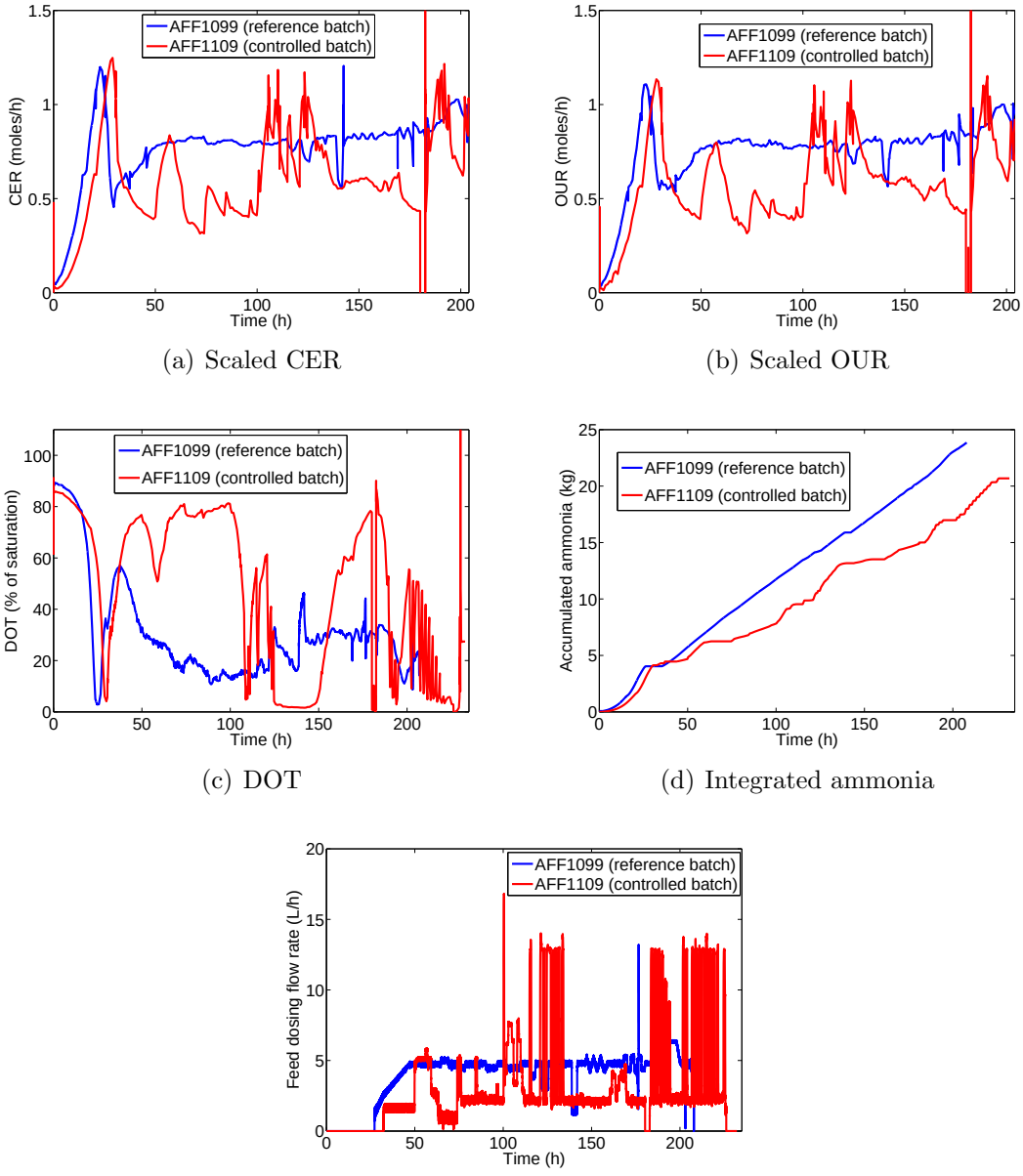
	Int. NH_3	DOT	CER	OUR
Int. NH_3	50	0	0	0
DOT	0	2	0	0
CER	0	0	2	0
OUR	0	0	0	2

The reason for the relatively large penalty on the integrated ammonia is that the order of magnitude is much larger than for the other three outputs. The value for R is chosen to 1 as the only input is the feed flow rate.

The batch used as reference for the GoLM framework and also used as the reference trajectory in the MPC controller is chosen to be AFF1099 which is a batch run without control and leading to a satisfactory enzyme production.

9.4.2 Evaluation of controlled cultivation

The variable trajectories of the controlled outputs obtained in the controlled cultivation experiment (AFF1109) are compared to the ones of the reference batch (AFF1099) in figure 9.6. Figure 9.7 shows the obtained variable trajectories for the variables which are not controlled.



(a) Scaled CER

(b) Scaled OUR

(c) DOT

(d) Integrated ammonia

(e) Feed dosing flow rate (input)

Figure 9.6. Comparison between the MPC controlled batch, AFF1109, and the batch used as reference for the controller, AFF1099

In general it is observed that most of the output trajectories are very oscillatory compared to standard behaviour (appendix A) which is due to an oscillatory input signal (feed flow rate) from the controller. The reason for this is that the Kalman

filter had a tendency to oscillate at some points during the process due to unexpected disturbances in the air flow rate (figure 9.7 (c)). The controller performs satisfactorily up until around 90h where the airflow suddenly decreases by around 5% and even stops completely at 111h for a few minutes. The airflow rate is not observed directly by the Kalman filter and therefore presumably affects the state estimation indirectly because it leads to a change in the DOT (figure 9.6 (c)). This fits the observation that the large oscillations in the input signal (figure 9.6 e) begin when the DOT starts to decrease rapidly (around 100h). Large oscillations are also observed in CER and OUR (figure 9.6 (a) and (b)).

Due to the unforeseen disturbance which is not handled by the current controller setup it is only reasonable to assess the performance of the controller until around 90 hours. It is observed that the DOT (figure 9.6 (c)) is significantly higher in the controlled batch than for the reference batch which is most likely connected to the feed dosing flow rate (figure 9.6 (e)) being considerably lower than for the reference during the first 100 hours of the cultivation. The CER and OUR (figure 9.6 (a) and (b)) have a similar evolution during the batch where the controller is still not active. CER and OUR drop slightly until around 50 hours where the controller increases the dosing feed flow rate. This change immediately increase CER and OUR and decrease the DOT significantly.

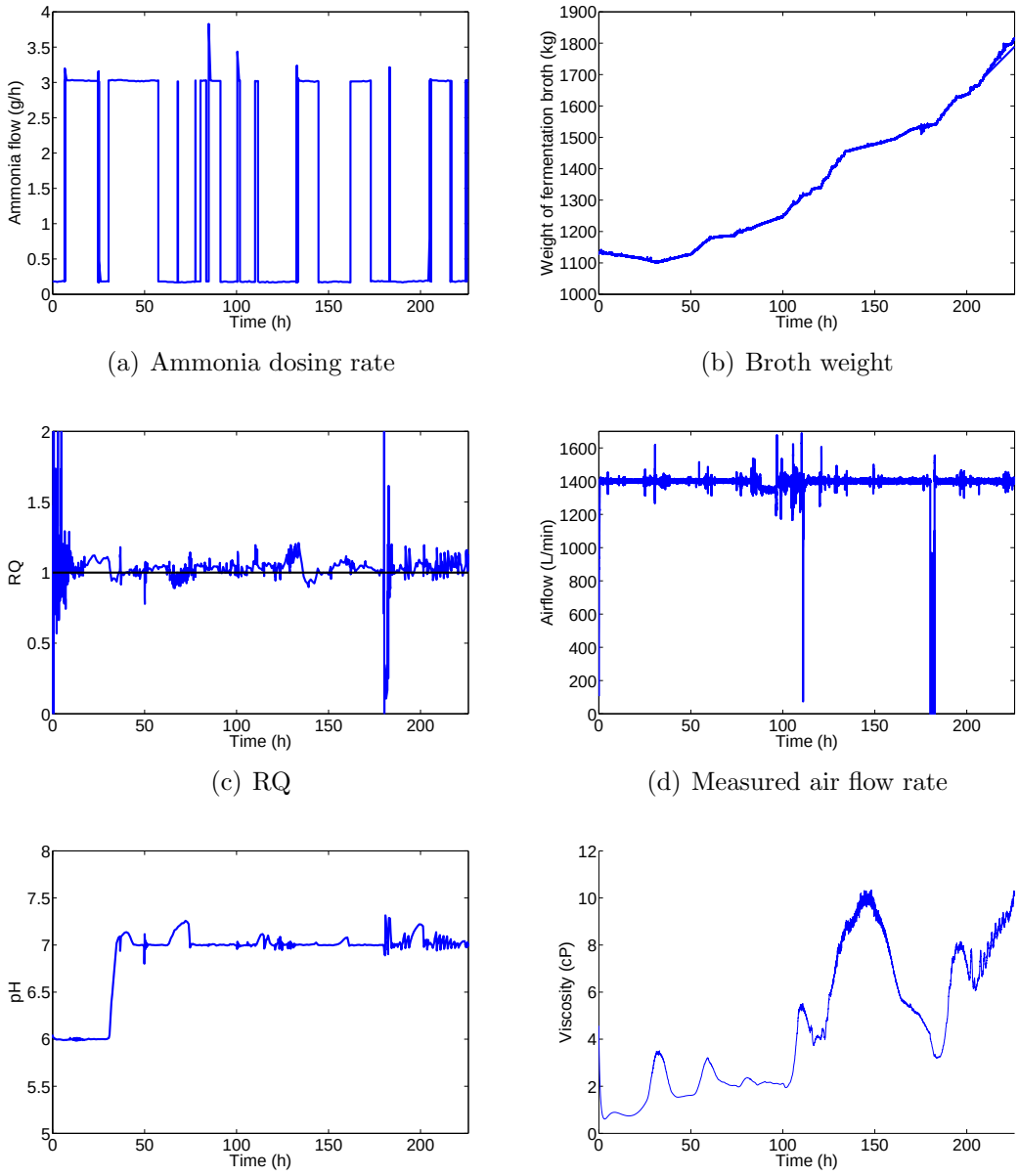
The oscillations in both input and outputs dampen gradually but at 180h the air flow is stopped completely for nearly three hours due to a failure in the central compressor system. For this period the cultivation has been proceeding without aeration which has introduced very large disturbances to the system. The Kalman filter was not able to stabilise after this disturbance.

The deviations observed in the output variables during this controlled cultivation experiment are far away from the standard operating region under which the GoLM model has been developed. The process variables have therefore not been within the range of validity dictated by the model and a good controller performance after large disturbances can therefore not be expected with the current controller setup.

In order to account for major external disturbances like the one observed either the model needs a training data set including batches experiencing this kind of disturbances or the operator needs to stop the Kalman filter and the controller when the stop in air flow rate occurs. When the air flow rate is back on track the operator can restore the operation of the Kalman filter and the controller. This functionality could be implemented in the process monitoring layer in the control hierarchy on figure 9.3.

Table 9.3. Biomass concentrations achieved in the MPC controlled batch, AFF1109

Time (h)	Biomass (g/L)
24	8.2
72	7.4
120	20.8
168	19.5
226	30.3



(a) Ammonia dosing rate

(b) Broth weight

(c) RQ

(d) Measured air flow rate

(e) pH

(f) Viscosity

Figure 9.7. Results of the MPC controlled batch, AFF1109

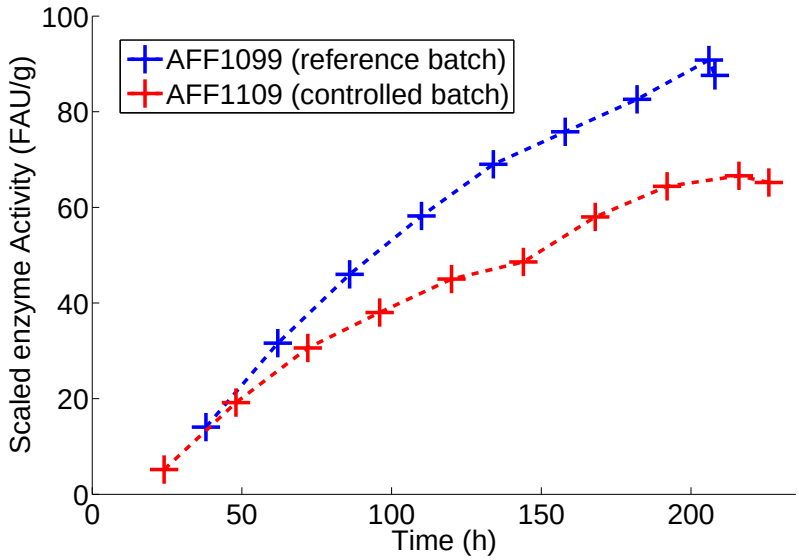


Figure 9.8. Enzyme activities obtained in the MPC controlled batch, AFF1109

Figure 9.8 gives the measured enzyme activities during the MPC controlled cultivation. The obtained enzyme activities are significantly lower than the ones obtained for the reference batch (AFF1099). In AFF1099 a final enzyme activity of 88FAU/g (scaled) was obtained compared to 65FAU/g in this cultivation. This is 26% less and can be explained by the large fluctuations in the input variable.

Due to technical difficulties it has not been possible to make precise biomass concentration measurements, the more reliable ones are shown in table 9.3. The results indicate that the biomass concentration is significantly lower than for a standard batch (appendix A). The final concentration is around 30 g/L whereas a standard batch can reach around twice that level.

A very interesting observation is that the viscosity (figure 9.7 (f)) decreases at around 140h which coincides with a significantly decreased feed flow rate. This observation supports the hypothesis made in chapter 6 that hyphae fragment under substrate starvation. This means that it is in fact possible to reduce the viscosity if problems with oxygen transfer are observed. The biomass does not seem to suffer too much due to the lack in substrate as the OUR and CER are back to the normal levels when the feed dosing is restored. The price for the decrease in viscosity is that only a marginal amount of product is formed during the period where viscosity is decreased. This finding can have some interesting implications for future control systems provided they become able to assess if the potentially improved oxygen transfer and thereby enzyme production rate makes up for the loss in productivity during the period needed to reduce the viscosity.

9.5 Conclusion

This chapter has introduced some perspectives on the introduction of automatic control in an industrial fed-batch cultivation. The GoLM model used in the controller has been introduced as well as the MPC control framework used.

Experience shows that the transition between batch and fed-batch mode is particularly challenging because it occurs at different points in time. For this reason the batch part has been truncated in the data sets used for modelling and only the fed-batch part has been modelled. This leads to a better performance of the model as the assumption behind GoLM stating that the process is time invariant from batch to batch can be better fulfilled.

The controller has been implemented at the Novozymes Fermentation Pilot Plant and a controlled cultivation has been carried out in order to evaluate the controller. Disturbances in the air flow rate occured during the process which lead to poor state estimations by the Kalman filter which ought to be turned off by the process operator. However this possibility was not included in this first implementation.

An important observation made was that it is possible to decrease viscosity by reducing the feed dosing rate significantly for a certain amount of time. This observation could prove beneficial for future control strategies.

10

Conclusions and future work

This section summarises the main results of this thesis and presents ideas for future work.

Fed-batch cultivations have major industrial importance but successful operation still relies on constant supervision by process operators. Current operation experiences large variabilities in the final product quality. Reducing the variability is highly desired because it allows for operation closer to the system constraints which improves overall process economy. Introduction of automated control schemes have the potential to reduce variability but are considered difficult to implement due to the limited availability of suitable mathematical models. A mathematical model of the process in question is necessary in order to implement advanced control schemes.

The aim of this thesis is to provide suggestions for development of such models based directly on data from industrial plants. This yields a very large potential as the amount of laboratory work can be reduced while more reproducible quality is obtained.

In this work the industrial cultivation of the filamentous fungus *Aspergillus oryzae* at Novozymes A/S has been studied. Experimental data as well as laboratory facilities have also been supplied by Novozymes A/S.

10.1 Data quality

Whenever process data are obtained some errors or inaccuracies will inevitably be introduced due to measurement principles, variations in the property measured, transmission noise, loss in the storage system etc. Therefore it is important to assess what type and degree of uncertainty applies to the data. In this work material balances have been applied to assess the quality of the data obtained from the Novozymes Fermentation Pilot Plant. In general these balances did not close very well, however the carbon balance was able to point out that the measured concentration of biomass was not reliable. It was also observed that other off-line measurements had problems.

The online measurements are generally regarded as reliable. It must be noted that some signals are based on measurement from more sensors, combining the uncertainty for each measurement. As an example the CER and OUR signals stem from a combination of the air flow sensor and a mass spectrometer.

The quality of the data used in this work is not worse than from any other industrial plant. It is just important to keep the measurement uncertainty in mind to avoid introduction of artificial dependencies and drawing conclusions on a wrong basis.

10.2 Black-box modelling

Black-box or data-driven modelling has been investigated for data sets from both an industrial plant and from pilot plant. The modelling framework used is proposed by Bonné and Jørgensen (2003) and is called Grid of Linear Models (GoLM) and the approach used is to divide the entire batch duration into a large number of grid points and fitting locally linear models to each of them. The combination of all these models is able to approximate the behaviour of the entire process which confirms the hypothesis that the data obtained is reproducible and contain sufficient information for development of predictive models (section 1.2).

Chapter 8 and 9 investigate the predictive capabilities of the GoLM modelling framework. In general the models developed possess good predictive capabilities but challenges are observed in the shift between batch and fed-batch phase. The reason is that the shift from batch to fed-batch occurs at different points in time and this is in contradiction to the major assumption in GoLM which states that the process is time varying within the batch but time invariant from batch to batch. This means that the same grid point model is not necessarily valid at the same time point for all batches, rather it is valid at a certain biological stage of the process. To circumvent this the data can be truncated after the batch phase has finished. In this way only the fed-batch part is modelled and the assumption of time invariant behaviour between batches can be fulfilled.

The data-driven models can only be reliably applied within the process window defined by the data used for the model estimation. This necessitates large excitations in the input variables in the data sets used for model estimation. Industrial processes however are generally run with exactly the same input trajectory from batch to batch in an attempt to minimise variability in the process. This means that industrial data in general do not provide very good information for estimating data-driven models. In this work the lack of excitation also has been observed. This problem can be partially circumvented by making small perturbations to the process as long they do not influence the resulting product quality. In this way large data sets with perturbed input signals could be generated in a relatively short time.

The structure of the GoLM framework makes it well suited for implementation in a controller framework. A controller framework based on GoLM has been presented and has been implemented in the Fermentation Pilot Plant and a controlled cultivation has been carried out to evaluate the performance of the controller. Large disturbances occurred during the process leading to poor state estimations which resulted in large fluctuations in the input signal and as a consequence the performance of the batch was below average. An important observation made was that it is possible to decrease viscosity by reducing the feed dosing rate significantly for a certain amount of time. This observation could prove beneficial for future control strategies in case the controller is able to assess if the potential increase in productivity makes up for the production lost in the period with reduced feed dosing rate.

10.3 Grey-box stochastic modelling

The grey-box stochastic modelling framework proposed in Kristensen *et al.* (2004) has been used in this work. The approach given in this framework is to combine first principle engineering models with operational data to produce predictive models suited for control purposes. The method combines a set of stochastic differential equations describing the dynamics of the system in continuous time with a set of discrete time measurements. An important advantage using this approach compared to using deterministic models is that that stochastic models can account for random variations in data. The stochastic differential equations include both measurement and process noise terms which makes it possible to distinguish noise from unmodelled dynamics.

The methodology has been used on the data available from the Fermentation Pilot Plant and has proven successful to determine functional relationships in data. The seed cultivation has been analysed with this framework (chapter 4) and it has been found that specific growth rate varies significantly during this period. In the beginning of the seed cultivation a lag period is observed and hereafter the growth rate increases drastically. The amount of biomass in the seed cultivation tank can be estimated based on on-line signals which means that the length of the batch phase in the main cultivation tank can be estimated as soon as the transfer of biomass is made. This allows for use of eg. different dosing feed policies which allow for better utilisation of the production equipment.

The results obtained for the main cultivation (chapter 5) show that at a certain biomass concentration the specific growth rate starts to decrease. This can be explained by a limitation occurring at this point, possibly due to formation of pellets. Pellet formation has been investigated in chapter 6 and it has been found that in fact large pellets are formed during the batch phase.

The analysis of the seed cultivation and batch part of the main cultivation confirms the hypothesis that the observed phenomena are reproducible and occur at the same biological progress of the cultivation (section 1.2) if the biomass concentration is used as a measure of the biological progress.

The concentration of initial starch in the cultivation tank has a large influence on the viscosity of the medium. This seems to lead to a mass transfer limitation for the batches run with standard initial starch concentrations. Batches having a low initial starch concentration did not experience this limitation and the substrate is also utilised more efficiently in these batches. The yield coefficients for conversion of substrate to biomass is higher for the batches with a low initial substrate concentration suggesting that a less efficient metabolism sets in at a certain biomass concentration.

The observations made on the available pilot plant data have led to the decision to carry out an additional experiment based on the standard recipe but having a much more detailed sample analysis than standard batches (chapter 6). This batch has contributed significantly to the understanding of many aspects of the cultivation. The experiment has been run according to a standard recipe except the last part of the process has been exposed to oxygen limitation. Frequent sampling has been used to identify the trajectories of biomass, enzyme concentration and metabolite concentrations.

The material balances fit relatively well considering the scale of the equipment thus the obtained data are relatively reliable. Two metabolites which were not observed before were identified, namely malate and mannitol. Malate is formed during periods with high glucose concentration which is believed to be a result of an overflow mechanism. Malate can substitute glucose during periods with insufficient glucose supply. Mannitol is formed under oxygen limitation as a product of substrate level phosphorylation and this compound is not observed to be metabolised. The highest conversion of substrate into product was obtained at low glucose concentrations and a DOT level above 6%. It was found that the process could be divided into six distinct operating regions, each possessing unique conditions regarding growth and product formation.

A particle size analysis was carried out which led to a number of important findings. Large pellets are transferred from the seed tank and later these large pellets are broken up due to substrate limitation and shear stress. The fragments produced have very small particle sizes and some time after the dosing feed has started a balance between pellet growth and break up develops. It was also observed that large particles, which are likely to be freely dispersed hyphae with long filaments, are formed during oxygen limitation. The change in morphology can be related to the viscosity. It has been observed that as the number of free hyphae increases the viscosity also increases.

Further analysis of the experimental data obtained in chapter 6 has been performed in chapter 7. This analysis led to proposal of a model for the entire cultivation based on the operating regions found in the experiment. The model has some serious shortcomings as only one type of biomass is assumed to exist. Therefore a more refined model containing both active and inactive biomass was proposed. This model gave a better description of the observed experimental results but still lacked to handle the transition between operating regions. The significantly improved performance of the model containing two types of biomass strongly supports the hypothesis that a significant amount of inactive biomass is present.

The models proposed can serve as monitoring tools and with further refinement function as software sensors for assessing the final amount of enzyme in the cultivation.

10.4 Outlook and further work

This thesis has shown that it is possible to extract valuable knowledge about a biological system from industrial data.

In order to optimise the productivity of the process it must be known which factor is limiting at each point of operation. The key to achieving higher yields is through identification and elimination of such bottlenecks. In practice however constraints apply with respect to process equipment and the micro-organism itself. One of the bigger challenges in identifying these bottlenecks is that the variables related to metabolic activity are not reliably measurable.

Introduction of online methods based on eg. infra-red spectroscopy might provide a means to monitor formation of metabolites and utilisation of glucose. The production of enzyme is inhibited by glucose which means that the glucose concentra-

tion needs to stay at a low level throughout the productive phase. This is currently achieved by a recipe which keeps the glucose concentration at a rather conservative level. Online estimation of the glucose concentration could prove valuable as it allows for operation closer to the optimal operation point with respect to glucose concentration. Measurement of formed metabolites could prove to be a valuable tool for monitoring the metabolic behaviour of the micro-organism. This could provide information on the current operating region and necessary precautions could be taken.

The division into different operating regions has proven successful and further refinement of this model structure will most likely be able to provide the accuracy needed for serving as the process model in an advanced controller. The modelling complexity can be taken to a much higher level than presented in this work. Complicated models based on flux analysis in the interior of the cell might provide answers to many question. Another more detailed modelling could include population balances where each type of metabolism and morphology can be accounted for. Common for these methodologies is that they require vast amounts of highly reliable data. In practice the model complexity is to a large extent dictated by the availability of data. This makes black-box modelling frameworks very well suited for modelling industrial processes.

Appendices

A

Pilot plant batches

A.1 Process description

The cultivation is initiated by inoculation of a seed tank with the desired strain. When the initial amount of substrate in the seed tank has been used the contents are transferred to the main fermentation tank. The inoculation process usually lasts for 43-48h. The actual substrate concentration is not measured during the cultivation so the criterion for transfer to the main fermenter is that the pH level has to decrease to below 5.3. The pH in the inocculation tank is around 6.2 most of the time and starts to decrease rapidly close to the point where the initial substrate has been consumed. The substrate in the inoculation tank consists of soya and sucrose in equal proportions. The main fermentation tank contains an initial amount of substrate which is composed of corn starch, magnesium sulfate, ammonium sulfate, potassium dihydrogen sulfate, potassium sulfate, trace metals and anti foaming agent. Before the transfer from the sedd tank the entire main fermenter with contents is sterilised. The weight of the main fermentation tank is adjusted to 1000kg by addition of steam.

When the criterion for transfer from the inocculation tank has been met 100kg is transferred to the main fermenter. The main cultivation process starts as soon as this transfer has occured. No feed is added in the first part of the process but the pH is maintained at 6 by adding ammonia. Without addition of ammonia the pH would decrease due to the consumption of ammonia and possible excretion of acids by the microorganism. The aeration is started at a set point of 750L/min and is ramped up to 1400L/min during the first 30 minutes. The last set point corresponds to full power and is kept during the rest of the cultivation. The air is supplied from a central compresser servicing several fermenters. This means that disturbances occur on the air flow when the load of the other fermenters is changed. Ammonia is added in gaseous form to the pipe transporting the incoming air. The air enters the bottom of the fermenter through a simple sparger (the air hits a round plate and is mixed with the liquid).

The mixing in the fermentor is maintained by three six-bladed Rushton turbines. The set point of the stirrer speed is constantly 225rpm. In some cases this set point can not be maintained due to high broth viscosity. During the batch phase the dissolved oxygen tension (DOT) drops very fast due to the rapid growth of the microorganism. The substrate concentration is very high during this phase and this leads to a high specific growth rate of biomass. When the initial amount of substrate is exhausted the pH starts to increase spontaneously because the microorganisms cease to take up ammnonia as ammonia uptake is associated to growth. At this point the DOT has dropped to around 5-10%. The criterion for the feed start is

that the pH must increase to above 6.3. Feeding is now started from the dosing tank at a rate of 1.4L/h and is ramped up to 2.8L/h during the first 5 hours and is ramped up further so it reaches a level of 5.2L/h 20 hours after the feeding has started. This level is kept during the entire cultivation unless the operators find it necessary to change it. The pH set point is increased from 6 to 7, 3 hours after the feeding has started. This pH level is maintained through ammonia addition during the rest of the cultivation.

The feed comes from a seperate dosing tank which is prepared before the main tank is inocculated. The feeding is added using a pulse/pause system, meaning that the feed flow rate is controlled by varying the number of feed pulses added to the fermenter in a given time. The pulse volume is 0.1L meaning that a feed rate of 5.2L/h is equal to 52 pulses/h. The dosing tank is loaded with typically 400kg of corn starch, α-amylase (Termamyl 120L) , sodium hydrogen phosphate dihydrate and anti foam agent. The starch is degraded by the enzyme at a pH of 7 and a temperature of 90°C for two hours. The dosing tank is then boiled at 123°C for 1 hour and the weight is adjusted to 1000kg by addition of steam. Finally a sample is taken from the dosing tank and a measurent called RTS (Refractometer dry weight), based on the refractive index is taken. The value read on the refractometer is in Brix units which is equal to the weight percent of sucrose in the sample. As the refractometer is calibrated on sucrose, the value read is only completely accurate if the starch is only broken down into sucrose. The general opinion at Novozymes A/S is that the RTS value is a good indication of the amount of carbohydrates that can be taken up by the microorganisms.

The expected value of RTS is 33% and the actual measured value is given to the DeltaV control system. The DeltaV control system uses this information to change the given feed flow set point in order to ensure that the desired amount of carbon is fed into the fermenter.

The fermentors are equipped with sensors for online measurements of different properties and these measurements are constantly monitored by process operators. If one or more of the monitored variables is outside the operating region specified in the recipe action is taken by the operators by manipulating one or more of the inputs to the process. A common example is that if the DOT falls below a certain value the feed rate is decreased by a given percentage.

Samples are taken from the fermentor at regular 12 hour intervals for laboratory analysis of enzyme activity, ammonia concentration and biomass concentration. These off-line measurements are not used for control purposes but for later evaluation of the batch. The concentration of enzyme is measured indirectly by measuring the enzymatic activity of the broth sample. The performance of the batch is evaluated as the enzyme activity of the broth at the end of the batch.

A.2 General observations in data

A closer look at the data from the Fermentation Pilot Plant reveals a number of interesting features. A number of general trends can be observed in the pilot plant data. For the batches run using the standard recipe (AFF1082-AFF1100) it is seen that the DOT drops from 100% in the beginning of the cultivation to around 10% at

the end of the batch. Feeding is started and the DOT rises and makes an overshoot and then becomes more or less stationary around 30%. The OUR and CER rise almost exponentially in the beginning of the batch phase and then drop when the initial substrate has been used up. The drop is not as large as one would expect from the fact that the substrate concentration is virtually zero at this point. This could indicate that the microorganisms metabolise another substrate. A look at the RQ shows that a value close to 1 is maintained in the beginning of the batch phase. After around 10 hours it increases to 1.1 and around the point of substrate depletion it decreases to around 0.9. Ammonia is added to the process as a nitrogen source and to keep pH constant, if no ammonia is added the pH drops. This indicates that either acid is produced or base is used.

According to Carlsen (1994) pellet growth should occur at the pH levels used in these cultivations (pH 6-7). This means that the interior of the pellet might be subject to oxygen limitation. Carlsen (1994) observed a considerable ethanol formation and this is likely to occur in the oxygen limited inner part of the pellet. When freely dispersed elements were grown no ethanol production was observed. On the other hand significant amounts of pyruvate were observed both under pellet growth and freely dispersed growth.

A.2.1 Effect of initial substrate concentration

The 7 pilot plant batches can be divided into two groups depending on the initial substrate concentration used. Batches AFF1082, AFF1098, AFF1099 and AFF1100 have been run with an initial substrate concentration corresponding to the value given in the standard recipe. Batches AFF1101, AFF1102 and AFF1103 have been run with an initial substrate concentration of only around 10% of the one given in the usual recipe. The reason for choosing the smaller amount of initial substrate is a result of the morphologically structured model proposed by Agger et al. In this model a relatively large part of the total biomass will consist of inactive hyphae which require substrate for maintenance and which contribute to increase the viscosity. This leads to reduced oxygen transfer and thereby a lower production rate. The aim is to reduce the fraction of this inactive biomass and have a larger fraction of the part producing the enzyme. In practice these 3 batches have led to a significantly lower final enzyme concentration than for the standard recipe.

A.3 Description of pilot plant batches

A total of 9 pilot plant batches have been investigated. The last two cultivations (AFF1108 and AFF1109) have been carried out after the initial investigation of data and are therefore not included in the preliminary modelling chapter. For all batches the standard online measurements described in chapter 3 have been recorded but due to technical difficulties a number of process variables have been recorded for all the batches. An overview of the available information is given in table A.1.

Table A.1. Variables specific for each batch

Variable	Type	1082	1098	1099	1100	1101	1102	1103	1108	1109
RI	Online	-	x	x	x	x	x	x	x	x
Viscosity	Online	x	-	-	-	x	x	x	x	x
Biomass	Offline	x	-	-	-	x	x	x	x	x
Ammonia conc.	Offline	x	x	-	x	-	x	x	x	x
DV		-	-	x	x	x	-	-	x	-
PI		x	x	x	x	x	x	x	x	x

A.3.0.1 Batches carried out in pilot plant

A number of batches have been carried out in pilot plant using exactly the same strain of *Aspergillus oryzae* as was used in the production data where the industrial BatchPro data come from. The batches are labelled AFF1082 through AFF1103. Some of the batches have been carried out with automatic control of the feed rate and some have been carried out using the flow rate given in the cultivation recipe. All inoculation tanks have been prepared in the same way but the initial amount of substrate in the main cultivation tanks has varied. Also the RTS value in the dosing tank has been varying. An overview of these parameters for the different cultivations is given in table A.2.

Table A.2. Cultivations carried out in pilot plant

Batch number	Amount of corn starch in main tank	Amount of corn starch in dosing tank	RTS in dosing tank	Type of control
AFF1082	45kg	400kg	36.3%	Manual
AFF1098	50kg	400kg	31.6%	Manual
AFF1099	50kg	400kg	38%	Manual
AFF1100	50kg	400kg	n.a.	Automatic
AFF1101	4kg	486kg	37.8%	Automatic
AFF1102	4kg	486kg	37.8%	Automatic
AFF1103	4kg	486kg	n.a.	Automatic
AFF1108	50kg	400kg	n.a.	Manual
AFF1109	50kg	400kg	n.a.	Manual

A.3.1 Estimation of density of dosing medium

The composition of the corn starch used for the dosing medium has properties which vary from batch to batch. The composition of the starch itself can change and also the content of water bound to the starch varies. The water content often lies around 10w/w%.

Table A.3. Cultivations carried out in pilot plant

Batch number	CTS	Enzyme activity	NH_3 concentration in tank	Viscosity	RI
AFF1082	12h	12h	24h	On-line	12h - Off-line
AFF1098	n.a.	24h	24h	On-line	On-line
AFF1099	n.a.	24h	24h	n.a.	On-line
AFF1100	n.a.	24h	24h	n.a.	On-line
AFF1101	12h	12h	n.a.	On-line	On-line
AFF1102	12h	12h	24h	On-line	On-line
AFF1103	12h	12h	24h	On-line	On-line
AFF1108	*	12h	24h	On-line	On-line
AFF1109	12h	12h	24h	On-line	On-line

*Measurements have been taken once an hour during batch and once every 8 hours during fed-batch

The starch supplied as initial substrate to the main fermenter is degraded by enzymes produced by the organism itself. The starch supplied to the dosing tank is degraded by enzymes added to the tank and the starch in this tank is fully hydrolysed before it is used in the main fermenter.

The actual density of the dosing medium has a very large effect on the mass balance and especially the carbon and degree of reduction balance as only the volumetric feed flow of the media is measured. The exact amount of carbon supplied to the main fermenter is therefore not known. Assuming that the carbon content per mass unit is always the same the actual amount of carbon added to the system is directly proportional to the density, so it is very important to have an accurate measure of this property. After preparation of the dosing tank a sample is taken and a refractometrical dry matter measurement (RTS) is made. This measurement is the Brix value of the medium and expresses the weight percentage of hydrolysed sugars. The apparatus is calibrated on sucrose but in practice this does not affect the result very much. Table A.4 shows the measured RTS value and the one calculated assuming a water content of 10w/w% in the starch and that all the starch can be hydrolysed into small suger units. In two cases the measured RTS value has not been written into the records and in two cases the final mass of the dosing tank is erroneous.

Table A.4 shows that the measured RTS value exceeds the calculated one suggesting that the water content is lower than the assumed 10w/w%.

In order to calculate the density of the dosing tank a simple mass balance is set up for the dosing tank. This can be used to estimate the density of the dosing medium because the total volume of withdrawn dosing is known. The volume withdrawn has been compared to the weight of the dosing tank and as expected a linear relationship between the two variables is found. The density and composition of the dosing tank is therefore assumed to be constant throughout the duration of the process. One could imagine that the bottom part is richer in carbon due to gravitation but due to rigorous stirring this seems not to be the case. Assuming constant density of the dosing medium the weight of the dosing tank can be expressed as:

Table A.4. Expected weight percentage in each batch

Batch	Mass of starch	Total mass	Calculated w/w%	Measured RTS
AFN1082	400	n.a.	27.7	36.3
AFN1098	400	1188	30.3	31.6
AFN1099	400	1000	36.0	38
AFN1100	400	1186	30.4	n.a.
AFN1101	486	1210	36.1	37.8
AFN1102	486	1230	35.6	37.8
AFN1103	486	487?	n.a.	n.a.
AFN1108	400	1080	33.3	34.3

$$M(t) = M_{t=0} - \rho \int_{t=0}^{t} F_{dos}(t)dt \tag{A.1}$$

Table A.5. Estimated density in each batch

Batch	Calc. dens. (kg/L)	95% conf. int. (10^{-4})
AFF1082	0.9808	35.0538
AFF1098	1.1325	12.0318
AFF1099	1.1773	5.5638
AFF1100	1.1532	7.4951
AFF1101	1.1924	8.3784
AFF1102	1.1942	5.6799
AFF1103	1.2176	82.1004
AFF1108	1.1813	5.4369

Table A.5 shows the calculated density and the 95% confidence interval. The density of the medium varies quite a lot from batch to batch so it is important to use the correct values for the mass and carbon balance calculations. The results indicate that there seems to be a correlation between RTS and density. Higher density is connnected to a higher RTS value and vice versa. When the RTS measurement is available it is a better measure of the carbon content than the calculated content because no assumptions about the amount of water are made. The property measured by the refractometer is refraction of light sent into the sample and this depends heavily on the concentration of carbon atoms with a chiral center. In the following it is assumed that only the carbon atoms are responsible for the refraction and that the brix measurement is proportional to the concentration of carbon atoms. Using these assumptions and taking into account that the refractometer is calibrated on sucrose the concentration of carbon atoms per weight unit can be calculated as:

$$\frac{n_c}{m_{total}} = \frac{[Brix]}{100}\frac{12}{M_{sucrose}} \tag{A.2}$$

Calculation of the concentration of carbon atoms per volume unit requires the density of the solution:

$$\frac{n_c}{V_{total}} = \rho \frac{[Brix]}{100} \frac{12}{M_{sucrose}} \tag{A.3}$$

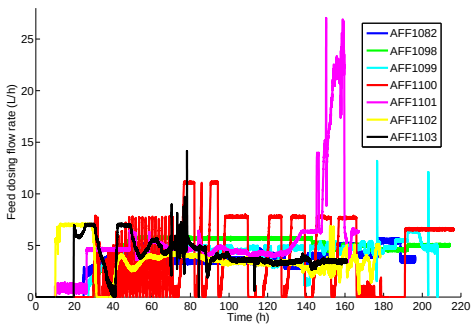
The calculated carbon content for each batch is given in table A.6

Table A.6. Estimated carbon content in each batch

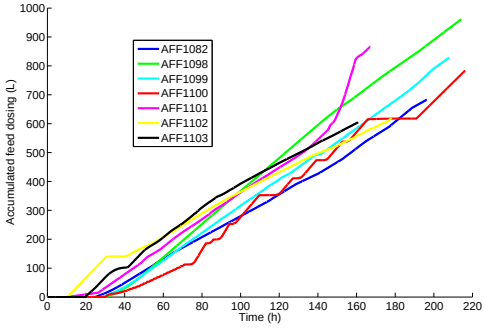
Batch	Calc. carbon content (C moles/kg)	Calc. carbon content (C moles/L)
AFF1082	12.73	12.48
AFF1098	11.08	12.55
AFF1099	13.32	15.68
AFF1100	NA	NA
AFF1101	13.25	15.80
AFF1102	13.25	15.82
AFF1103	NA	NA
AFF1108	12.02	14.20

A.3.1.1 On-line data for pilot plant batches

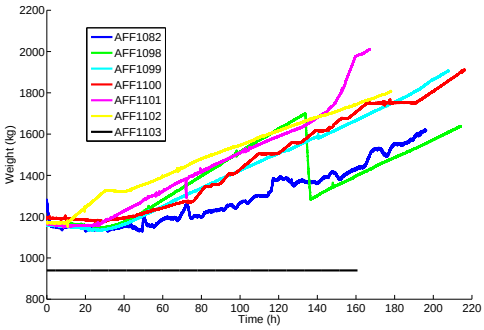
In the following the most important on-line data are shown for 7 batches between AFF1082 and AFF1103.



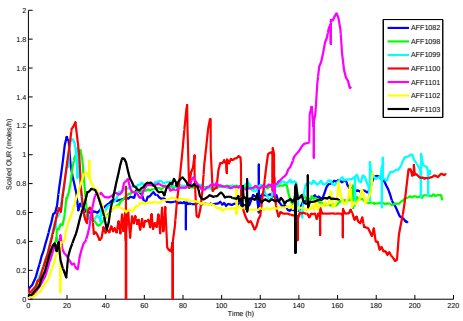
(a) Measured dosing feed rate



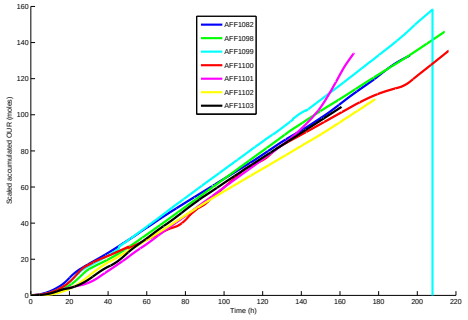
(b) Accumulated dosing feed



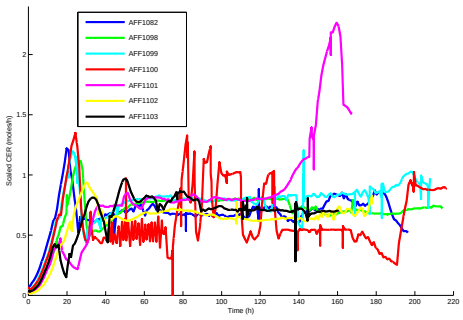
(c) Broth weight



(d) Scaled OUR

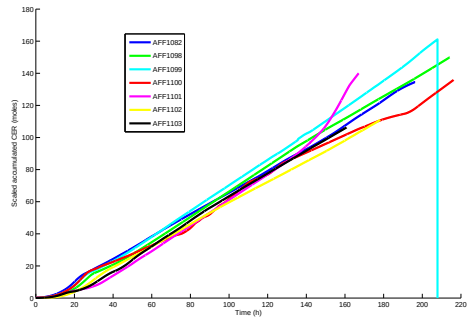


(e) Scaled integrated OUR

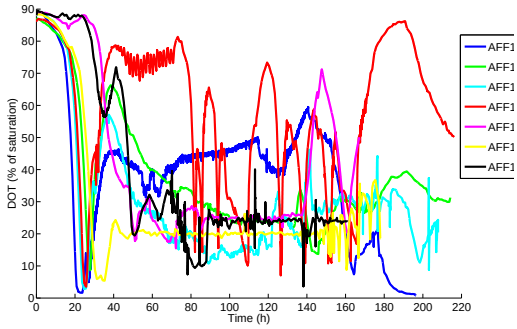


(f) Scaled CER

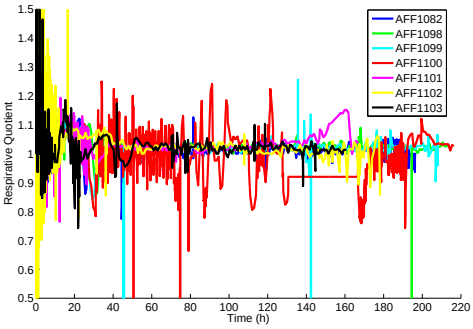
Figure A.1. Overview of 7 pilot batches



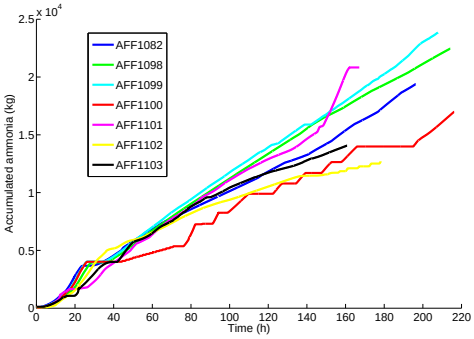
(a) Scaled integrated CER



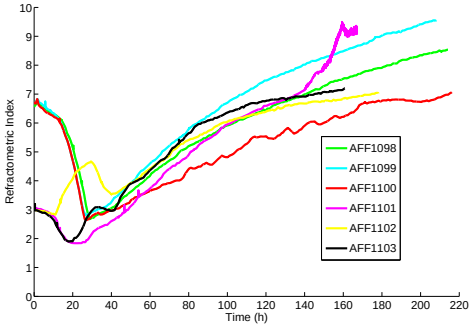
(b) DOT



(c) RQ



(d) Integrated ammonia dosing



(e) Refractive index

Figure A.2. Overview of 7 pilot batches

The batches have been run using different strategies and can be divided into different categories. Some batches have been run with automatic control and some batches have had a reduced initial amount of starch. The most important features of each of the batches is given in the following:

- AFF1082:
  This batch has been carried prior to the beginning of this project. The batch is run according to recipe except that perturbations have been introduced in the dosing feed rate in order to investigate the influence on DOT.

- AFF1098:
  This batch is the first one carried out in this project. The purpose was to run a batch according to recipe and compare the outcome to industrial scale data. At 138h the sampling valve did not close properly and almost 400kg of the fermentation broth was lost. This lead to overdosing in a period as the fault was not observed immediately.

- AFF1099:
  This batch has also been carried in this project and has been run according to recipe. Perturbations have been introduced in the feed dosing rate in order to investigate the input/output behaviour of the system and provide data for the data-driven modelling.

- AFF1100:
  This batch has been run under another project in order to investigate the performance of a control based strategy. The control strategy was to follow an optimal substrate concentration obtained from a model based estimator. The controller had problems with stability and large variations in the feed dosing rate occured. This is reflected in process variables like DOT, OUR and CER.

- AFF1101:
  Again automatic control has been used but the amount of initial starch has been reduced to around 10% of the usual value as an attempt to reduce the viscosity. The control strategy was to keep DOT constant at 25% to avoid oxygen limitation. This went relatively well up to 137h where the controller became unstable. The flow increased dramatically leading to a very high viscosity towards the end of the batch.

- AFF1102:
  This batch is similar to AFF1101, the control strategy was to keep DOT at 20% and the performance was significantly better than in the previous batch. At 140h the controller became unstable again and the feed dosing fluctuated a lot leading to fluctuations in DOT.

- AFF1103:
  This batch has been run in a way similar to the 2 previous ones. The strategy was to keep DOT at 25% but this went less well than for AFF1102. The weight measurement has been faulty in this batch. The weight always stays constant at 939.5kg and is useless for any purposes.

A.3.1.2 Off-line data for pilot plant batches

Table A.7. Off-line data for AFF1082

Time (h)	EnzA (FAU/g)*	CTS (g/L)	NH_3 conc. (mg/L)
7	0.07	6.9	conf.
19	4.3	9.2	conf.
31	10.9	18.5	conf.
43	22.4	-	conf.
55	29.4	17.7	conf.
67	36.8	18.8	conf.
79	43.4	20	conf.
91	49.8	21.2	conf.
103	54.4	33.5	conf.
115	61.2	31.2	conf.
127	58.2	26	conf.
139	63.2	29	conf.
151	66.4	24.6	conf.
163	72.2	26.6	conf.
175	80.2	33.5	conf.
187	82.0	57.9	conf.
197	88.8	61.5	conf.

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.8. Off-line data for AFF1098

Time (h)	EnzA (FAU/g)*	NH_3 conc. (mg/L)
40	15.1	conf.
64	29.4	conf.
88	40.8	conf.
112	52.2	conf.
136	57.6	conf.
160	68.2	conf.
184	73.6	conf.
208	72.4	conf.
217	72.8	conf.

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.9. Off-line data for AFF1099

Time (h)	EnzA (FAU/g)*	NH_3 conc. (mg/L)
38	14.0	conf.
62	31.6	conf.
86	46.0	conf.
110	58.2	conf.
134	69.0	conf.
158	75.8	conf.
182	82.6	conf.
206	90.8	conf.
208	87.6	conf.

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.10. Off-line data for AFF1100

Time (h)	EnzA (FAU/g)*	NH_3 conc. (mg/L)
43	15.4	conf.
67	24.8	conf.
91	30.6	conf.
115	40.2	conf.
139	45.8	conf.
163	51.6	conf.
187	51.2	conf.
211	55.6	conf.
216	51.2	conf.

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.11. Off-line data for AFF1101

Time (h)	EnzA (FAU/g)*	CTS (g/L)
23	2.3	-
35	11.1	7.5
47	12.4	10.1
59	18.5	15
71	26.0	18.3
83	39.2	18
95	40.0	20
107	44.8	22.6
119	50.6	22.3
131	-	27.2
143	65.0	31.8
155	57.0	40.1
167	55.2	58.4
167	51.2	-

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.12. Off-line data for AFF1102

Time (h)	EnzA (FAU/g)*	CTS (g/L)	NH_3 conc. (mg/L)
23	1.7	6.7	conf.
35	8.6	13.5	conf.
47	14.0	13.5	conf.
59	20.4	14.1	conf.
71	25.0	13	conf.
83	30.4	13	conf.
95	36.6	15.2	conf.
107	39.0	16.7	conf.
119	41.0	19.2	conf.
131	46.6	25.9	conf.
143	44.8	22.3	conf.
155	-	23.8	conf.
167	45.6	26	conf.
179	-	33.1	conf.
180	45.0	23.8	conf.
180	45.6	-	conf.

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.13. Off-line data for AFF1103

Time (h)	EnzA (FAU/g)*	CTS (g/L)
25	2.9	6.6
37	7.9	9.4
49	11.9	10.4
61	19.7	9.7
73	24.8	12.8
85	30.0	14.2
97	35.2	16.3
109	39.2	19.2
121	43.8	-
133	46.2	-
145	48.6	-
157	50.4	-
161	-	-

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.14. Off-line data for AFF1108

Time (h)	EnzA (FAU/g)*	CTS (g/L)	NH_3 conc. (mg/L)
18.08	-	-	-
29	-	-	27.2
38.17	-	-	-
41	13.7	-	-
53	19.7	17.5	850
65	29.4	-	-
77	36.4	22.8	909
89	43.4	-	-
101	51.6	22.8	876
110	-	-	-
113	55.2	-	-
125	59.8	31.1	812
137	64.4	-	-
149	66.6	32.9	738
150	-	-	-
161	69.2	-	-
173	67.4	34	1136
185	70.0	-	-
197	65.0	37.3	1223
202	70.2	-	1254
203	-	-	-

*Due to confidentiality the enzyme yields are scaled by a constant factor

Table A.15. Additional off-line data for AFF1108

Time (h)	glucose (g/L)	malate (g/L)	mannitol (g/L)	CTS(fungal) (g/L)
6.33	3.1306	0.2895	-	-
7.33	3.1436	0.3149	-	-
8.33	3.2892	0.4658	-	-
9.42	3.1415	0.5795	-	-
10.50	3.1619	0.7500	-	-
11.33	3.0372	0.9126	-	-
12.58	2.8692	1.1528	-	-
13.67	2.7621	1.4964	-	-
14.58	2.6853	1.7349	-	-
15.50	2.7711	2.0217	-	-
16.50	3.2950	2.1801	0.4430	-
18.08	3.0466	2.4724	0.3370	5.37
19.08	2.6084	2.6609	-	-
20.08	2.3362	2.7916	-	-
21.17	2.0673	3.2518	-	-
21.92	4.4859	2.9256	-	-
31.08	4.8498	3.9473	-	-
32.00	3.6491	3.5283	-	-
33.00	2.1329	3.4402	-	-
34.17	0.2211	3.1309	-	-
35.00	0.0327	2.9970	0.2160	-
36.00	0.0171	1.7763	-	-
37.00	0.1050	1.4311	-	-
38.17	0.1453	0.9788	-	15.54
39.08	0.1159	0.7422	-	-
40.08	0.1506	0.1364	-	-
41.00	0.1445	0.0517	0.3620	-
45.50	0.0676	0.0586	-	-
55.17	0.1476	0.1396	0.4650	-
62.00	0.0571	0.1882	-	-
69.50	0.0220	0.2073	-	-
79.33	0.0081	0.1621	-	-
86.00	0.0352	0.1431	-	-
93.50	0.0792	0.0891	-	-
106.00	0.0833	0.0490	0.2970	-
110.00	0.0266	0.1932	-	19.10
117.50	0.0021	0.2120	-	-
126.00	0.0084	0.2193	-	-
133.00	0.0123	0.2233	-	-
141.50	0.0146	0.2462	-	-
150.00	0.0016	0.2630	-	23.59
158.00	0.1130	0.5171	-	-
165.50	0.1689	0.9186	1.6540	-
174.00	0.1754	1.1070	-	-
182.00	0.1843	1.1906	-	-
189.50	0.2603	1.2767	-	-
198.00	0.3117	1.3124	-	-
201.00	0.3878	1.2847	4.5700	-
202.00	0.3147	1.2716	-	-
204.00	0.2933	1.4368	4.4500	26.30

A.3.2 Compositions of media

The seed cultivation tank have had the same composition for all 9 batches but the concentration of corn starch has been varying for the main cultivation and dosing tanks.

Due to confidentiality the exact compositions of the media can not be revealed.

Table A.16. Concentration of corn starch for the main cultivation and dosing tanks

Batch	Main tank conc. (g/kg)	Dosing tank conc. (g/kg)
AFF1082	50	400
AFF1098	50	400
AFF1099	50	400
AFF1100	50	400
AFF1101	4	400
AFF1102	4	400
AFF1103	4	400
AFF1108	50	400
AFF1108	50	400

B

Materials and methods

B.1 Pilot plant instrumentation

The signals obtained from the DeltaV control system in the Fermentation Pilot Plant at Novozymes are briefly described in the following.

B.1.0.1 Cultivation time (h)

The cultivation time indicates the time in hours since the start of the cultivation. The starting point is defined to be when the contents of the seed tank are transferred to the main cultivation tank.

B.1.0.2 NH_3 flow rate (g/h)

NH_3 is provided as a means of keeping the pH constant and to supply nitrogen to the cultivation. A SISO controller (PI control) adds NH_3 in order to keep pH at the predefined value. The performance of this control loop is considered satisfactory.

B.1.0.3 Integrated NH_3 (kg)

This signal is the integrated ammonia flow rate signal. Manual integration verifies that the automatic integrataion made by DeltaV is satisfactory.

B.1.0.4 Dissolved O_2 tension (Percentage of saturation)

The DOT is measured with a sensor mounted in the bottom of the tank close to the lowest Rushton turbine. Before each cultivation the sensor is calibrated to a value close to 100% (the actual value is recorded by the operator). The liquid in the tank during calibration contains the initial amount of substrate at a pressure of 1.3barg. The tank is stirred and aerated hence total saturation can be assumed. After inoculation the pressure is reduced to around 1 barg and the DOT does not read 100% in the beginning of the actual cultivation because of the lower saturation concentration under these circumstances. The calculation of k_La must therefore be corrected by taking the actual pressure into consideration. Assuming that the oxygen transfer rate from gas to liquid is equal to the oxygen uptake rate (ie. no accumulation of gas in the cultivation tank), k_la can be calculated based on the measured oxygen uptake rate:

$$k_la = \frac{OUR}{V_l\left(c_l^* - c_l\right)} \tag{B.1}$$

V_l is the liquid volume in the cultivation tank, ie. the total volume of the broth minus the volume of gas bubbles. c_l^* denotes the saturation concentration under the actual conditions and c_l denotes the actual concentration in the cultivation tank. The saturation concentration of oxygen in the liquid is dependent on the actual partial pressure of oxygen. According to Henrys law the saturation concentration is proportional to partial pressure giving:

$$c_l^* = c_{l,p_{ref}}^* \frac{p}{p_{ref}} \tag{B.2}$$

The DOT sensor measures the actual oxygen concentration as a percentage of the oxygen concentration which was present during calibration. It is assumed that the liquid was saturated with oxygen at the calibration. This can be expressed as:

$$c_l = c_{l,cal}^* \frac{DOT}{DOT_{cal}} \tag{B.3}$$

DOT is the actual sensor reading and DOT_{cal} is the reading obtained under calibration (close to 100%). $c_{l,cal}^*$ is the oxygen saturation concentration during the calibration. Using Henrys law this can be found as:

$$c_{l,cal}^* = c_{l,ref}^* \frac{p_{cal}}{p_{ref}} \tag{B.4}$$

$$k_l a = \frac{OUR}{V_l c_{l,ref}^* \left(\frac{p}{p_{ref}} - \frac{pO_2 p_{cal}}{pO_{2,cal} p_{ref}} \right)} \tag{B.5}$$

Here V_l denotes the liquid volume and c_l^* denotes the saturation concentration of oxygen under the the reference pressure p_{ref}. p is the actual oxygen partial pressure and $p_{initial}$ is the oxygen partial pressure under which the DOT sensor has been calibrated. pO_2 and $pO_{2,initial}$ are the DOT readings at the actual time and at the time where the sensor was calibrated respectively.

B.1.0.5 pH (No unit)

The pH is measured with a pH electrode installed in the lower region of the cultivation tank. Before each cultivation it is calibrated using buffer solutions with pH values of 4 and 7 respectively. Samples are taken from the broth during the cultivation and the pH is measured in the laboratory. The result is available shortly after the sample has been taken and is recorded in the log book. If the on-line pH measurents deviates more than 0.2 pH units from the off-line measurement the on-line measurement is recalibrated to the value measured in the sample. This occurs rarely and it can be seen as a sudden jump in the on-line pH measurement. Noise: The measurement has some noise with a high frequency and low amplitude. The signal is smoothened out because there is no immediate physical reason that the pH should change so rapidly. It is more likely that it is due to varying local compositions (process noise) or signal noise.

B.1.0.6 Feed flow rate measured (L/h)

This variable is the actually measured volumetric feed flow to the tank. It deviates from the setpoint value if the RTS in the dosing tank deviates from the expected value. This signal seems to be very noisy but a closer look reveals that the noise comes from the opening and closing of the dosing valve. This signal is not used in the modelling, instead the feed flow rate set point is used.

B.1.0.7 Accumulated feed flow (L)

The accumulated feed flow is calculated by the DeltaV control system based on the setpoint of the feed flow and not on the actually measured feed flow. This signal has a staircase shape because of the pulse/pause feeding used. The signal is left unfiltered.

B.1.0.8 Feed flow set point (L/h)

The setpoint of the feed flow is the actual feed flow rate added to the tank with the corrected value of RTS. It has been confirmed that integration of this signal gives the accumulated feed flow. This signal is free from noise. So no filtering has been applied.

B.1.0.9 Back pressure (barg)

The back pressure is the pressure measured in the top of the tank, above the cultivation broth. In the industrial size cultivation tanks a sensor also measures the pressure in the bottom of the tank but this measurement is not available in pilot plant.

B.1.0.10 O_2 uptake rate - OUR (moles/h)

This variable is calculated by the DeltaV control system based on a mass balance on oxygen. The gas composition is measured on a central mass spectrometer. Prior to measurement the air is dryed, so all gas compositions correspond to dry air. The OUR is calculated as:

$$OUR = \frac{v_{air,in} p_{atm}}{RT}(y_{O_2,in} - y_{O_2,out}) \tag{B.6}$$

B.1.0.11 Accumulated O_2 (moles)

OUR is constantly integrated by the control system to give the accumulated value. It has been confirmed that integration of the OUR gives this signal.

B.1.0.12 CO_2 evolution rate - CER (moles/h)

This variable is calculated by the DeltaV control system based on measurents of the CO_2 content of the exhaust gas and the airflow rate of the cultivation tank. The CER is calculated as:

$$CER = \frac{v_{air,in} p_{atm}}{RT} y_{CO_2,out} \tag{B.7}$$

B.1.0.13 Accumulated CO_2 evolution (moles)

The same applies as for the accumulated OUR.

B.1.0.14 Weight (kg)

The weigth is measured by a sensor mounted right beneath the tank. The cultivation tank actually stands on the sensor. A large amount of high frequent noise is present in this signal. A hard filtering is used to smoothen out the noise. There is no physical reason that the weight should change rapidly. Some of the noise is due to vibrations from the stirrer. A number of pipes are connected to the cultivation tank and these can to some extent prevent counteract the free movement of the cultivation tank in the vertical direction.

B.1.0.15 Viscosity (cP)

The viscosity is measured on-line by a viscosity meter. The unit is centi Poise. Currently this measurement is only available in one of the tanks in pilot plant and in none of the tanks in the production facility. This signal is very noisy so a relatively hard filtering is applied. There is no physical reason that the viscosity should change that rapidly unless if different local compositions are measured.

B.1.0.16 Respirative quotient (RQ) (No unit)

This is the ratio between CER and OUR calculated on-line by the DeltaV control system.

B.1.0.17 Temperature (°C)

The temperature of the cultivation broth is monitored by a sensor located in the bottom of the tank. The temperature is controlled by cooling water running in spirals inside the cultivation tank. The control is quite efficient and no cooling problems are observed with this strain. There is some noise on this signal but the amplitude is very small. Typically around 0.05K. The signal is smoothened out because the temperature can not change rapidly because of the large heat capacity of the system. The temperature is effectively kept constant at 32°C.

B.1.0.18 Air flow rate (Nm^3/h)

This signal is the volumetric air flow measured just before the entrance to the cultivation tank. The unit is NL/min which means that the value corresponds to a temperature of 0°Cand an absolute pressure of 1 atm. The airflow out from the tank is not measured but will differ from the flow into the tank because of oxygen uptake, carbon dioxide release and water evaporation. Assuming $RQ = 1$ and that no accumulation of gasses occurs involves that the molar (or volumetric) flow rate

of dry air is the same in the inlet and outlet. However the mass flow rate changes due to the difference in molar weights of O_2 and CO_2. A central pump supplies pressurised air to all the cultivation tanks in the plant which means that the air flow can change depending on the load of the other cultivation tanks. Even though the air flow rate is set to a constant rate deviations are observed.

B.2 DeltaV and PI system

DeltaV is an advanced process control system made by Emerson Process Management. More information on DeltaV can be found on http://www.easydeltav.com/. A seperate Windows PC running Matlab has been set up in the control room. An OPC interface supplied by IPCOS (http://www.ipcos.com) has been used for communication between Matlab and the DeltaV control system. OPC stands for OLE for Process Control (see http://www.osisoft.com for more information). This set-up makes it possible to have two-way communication with the DeltaV system through Matlab scripts. This is a very convenient way to perform data collection and send inputs to the process. PI (Plant Information) is a system for data collection and database storage.

B.3 Data compression in PI

Prior to storage data is compressed by the PI system. Compression of data is a complicated trade-off between needed accuracy and available storage room. The allowable degree of compression depends on the purpose of the data. For visual inspection of data or identification of gross errors strongly compressed data will often be satisfactory. If the purpose is modelling fast dynamics of the process usually less compressed data are needed. Compression can introduce artifacts in the data and should be used carefully Watson *et al.* (1998).

The raw data from the DeltaV control system are compressed before storage in the database. The type of compression is sliding door?. The principle behind this compression system is that as long the measurement follow the same linear trend (ie. the slope of the signal is constant) data points are only stored with a specified rate. When the trend changes points are stored when the window no longer covers the current trend. Therefore a big difference in the sampling interval is often observed. The are both benefits and drawbacks of such a compression system.

The benefits are that a very large compression can be achieved and the periods where large changes occur contain relatively many stored data points.

The drawbacks are that information with high frequencies are lost (which also applies for most other compression techniques) but also that low frequency information becomes biased. This is shown in (ref). This means that the average value for a given time period can not be reproduced. This is not a big problem when the data is only used for giving an overview of the evolution of the batch but for modelling purposes it is a major issue.

B.4 Biomass measurement

Measurement of biomass concentration in filamentous fungi fermentations is generally very difficult and trustworthy results are hard to obtain. At Novozymes the biomass is estimated as the content of dry matter in the sample. This measurement includes all matter which can not be dissolved and therefore undissolved salts and impurities from the medium are also included in the measurement. The value is thus not very reliable and should be used with caution. Each sample has been measured twice to reduce the uncertainty and estimate the variation. One could argue that the true value of the biomass concentration is definetely not higher than the measurement. Two methods for estimating the biomass concentration are used. One method is referred to as the general dry matter measurement and is based on sequential centrifugation and washing of the broth. The other method is referred to as the fungal dry matter measurent and here the broth sample is filtered through a syringe where the piston has been taken out and a filter paper and approximately 1cm of small glass balls have been placed in the bottom. The first method is the fastest and is usually used even for fungal fermentations at Novozymes A/S. The latter method gives a better washout of impurities but is more laboursome. For the experiments carried out in this work the latter method has been used. The result is reported as mass of dry matter per volume of broth (g/mL).

B.5 Enzyme activity measurement

Broth samples are sent out from the house for activity analysis. The test method is based on measuring the amount of starch which can be hydrolysed by a given weight of broth. The result for the enzyme activity is reported in number of activity units per mass of broth sample (FAU/g) The FAU (Fungal Activity Unit) is an arbitrary unit and corresponds to 0.105mg of α-amylase.

B.6 Composition of biomass and enzyme

The exact biomass composition is not known but throughout this work it has been assumed that the composition given in Carlsen (1994) can be used. The composition here is with an ash content of 7.5w/w%. The sulphur content is not included in the given composition. The amount is most likely included in the ash content. Fortl om svovlindhold - referer. The enzyme composition is found in Carlsen (1994).

B.7 HPLC

A Water-Summit column has been used in the first HPLC analysis. 2 different standards have been used. A standard acid standard and a fungal standard. Each sample has been filtered in a Chromafix C18 cartridge from Macherey-Nagel. The eluent is sulphuric acid. UV and RI detectors have been used, so two spectra for each sample are recorded. A total of 59 samples have been run in this column.

The first HPLC analysis indicates that important compunds have not been taken

into account. Some of these are likely to be polyols. In order to investigate this further a second HPLC analysis has been carried out on a Dionex column. This column is run with a slightly basic eluent and it is well suited for seperating polyols. The standard used contains... 10 samples have been run with this equipment and it was found that especially the concentration of mannitol was much higher than in the samples.

B.7.1 Standard compositions

Two HPLC standards have been used to obtain a picture of which compounds are present in the fermentation broth. One standard contains mostly organic acids (the acid standard) and the other standard contains compounds which are usually found in fungal fermentation brohts. The composition of the two samples is given in tables B.1 and B.2. Each of the standards is diluted in 6 different dilutions to be able to obtain standard curves with 6 points for each compound.

Table B.1. Composition of acid standard

Compound	Concentration (g/L)
Glucose	20
Pyruvate	0.5
Succinate	0.5
Glycerol	1
Acetate	2
Ethanol	15

Table B.2. Composition of fungal standard

Compound	Concentration (g/L)
Succinate	1
Fumarate	0.1
Malate	0.5
Citrate	1
Oxalate	1

Table B.3. Composition of polyol standard

Compound	Concentration (g/L)
Arabitol	0.1004
Xylitol	0.098
Glycerol	0.2305
Erythriol	0.1135
Mannitol	0.0993

C

Particle size distribution data

C.1 Particle size distribution data

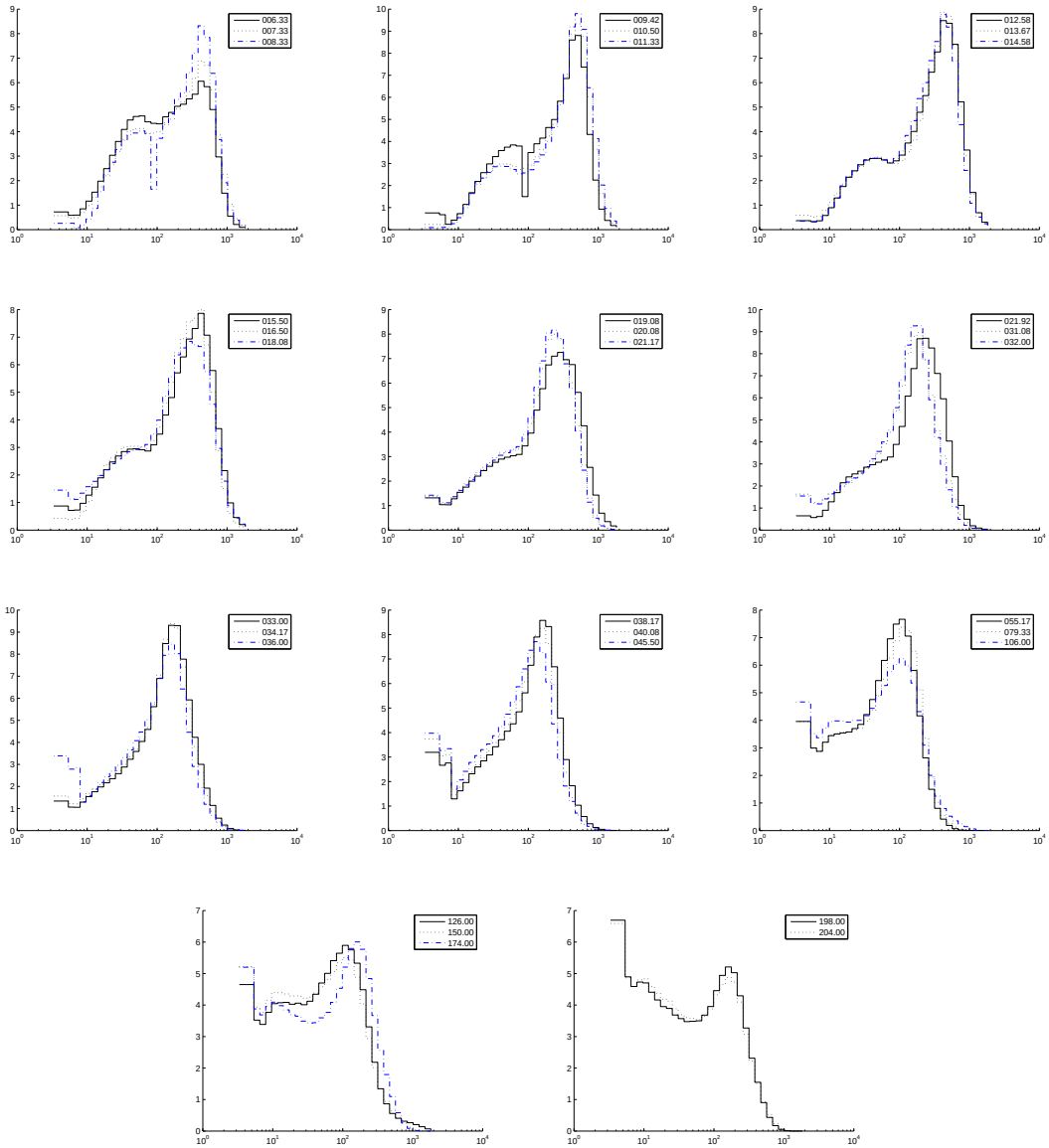


Figure C.1. Particle size distributions, the values are given in percent of total. x-axis: Particle diameters in μm. y-axis: Distribution in percent

C.2 Concentration distribution data

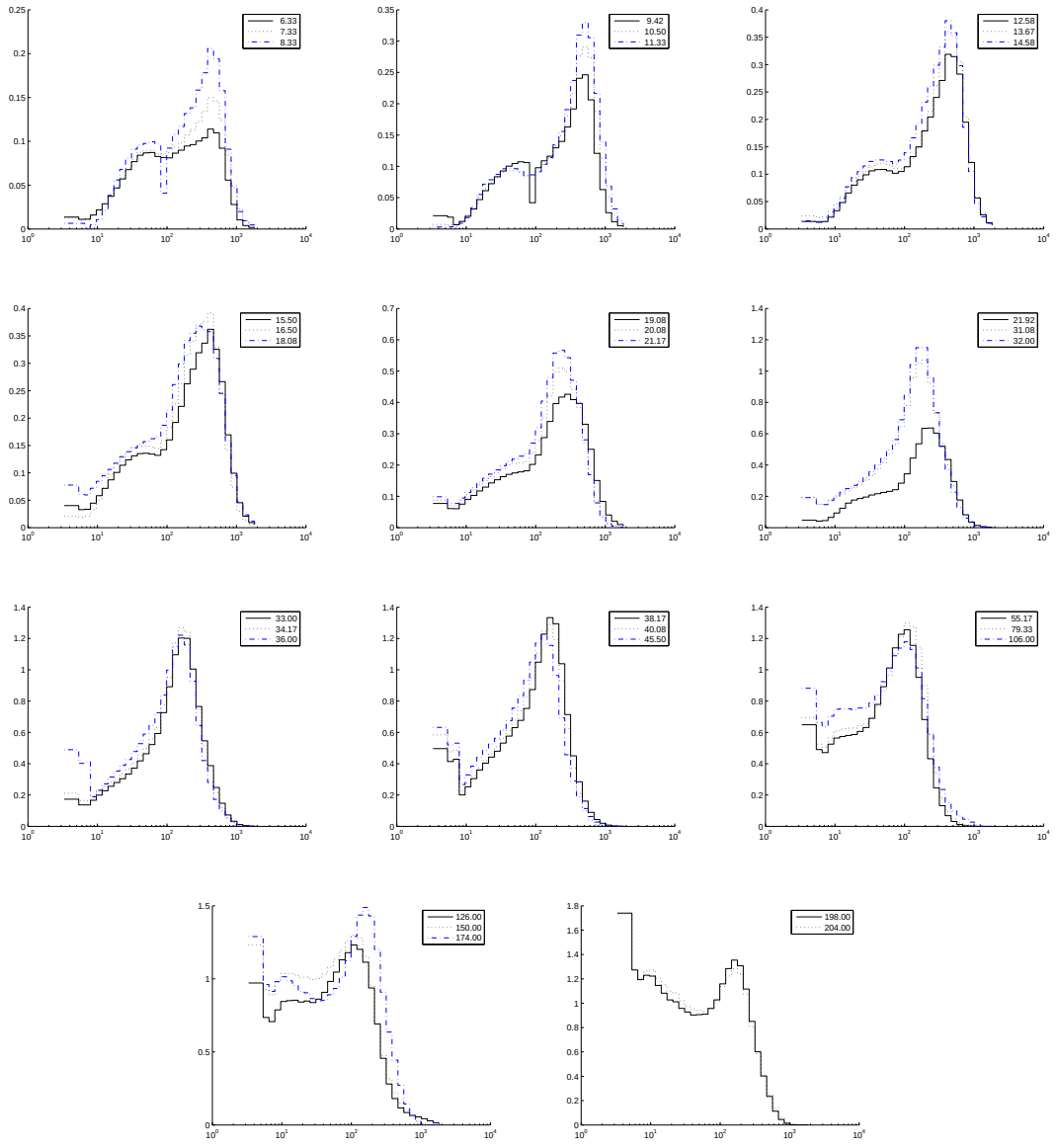


Figure C.2. Particle size distribtutions, given as concentrations. x-axis: Particle diameters in μm. y-axis: Concentration in g/L

C.3 Mass distribution data

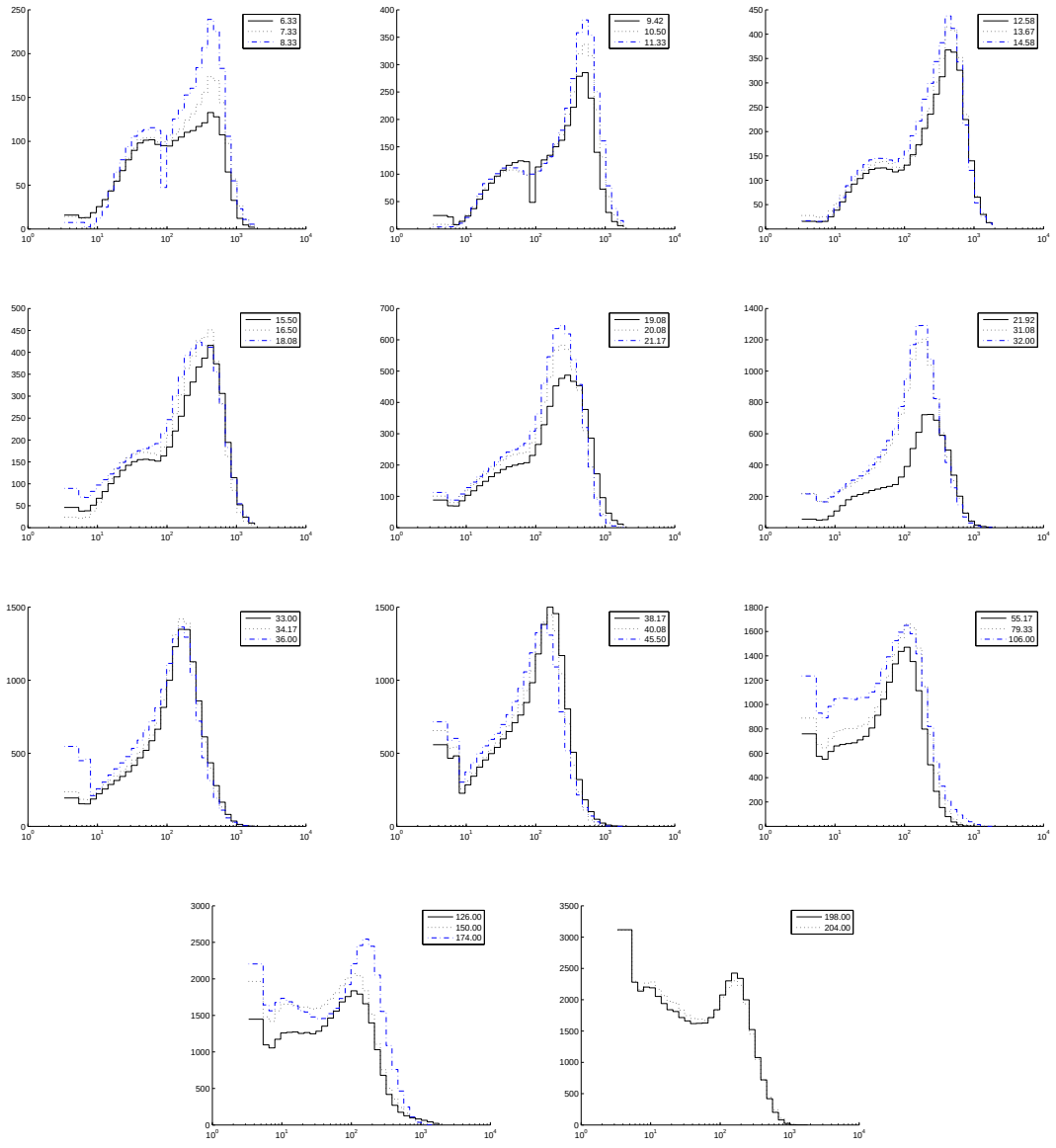


Figure C.3. Particle size distribtutions, given as mass within each size class. x-axis: Particle diameters in μm. y-axis: Mass in g

C.4 Calculated mean distributions

Table C.1. Particle size distribution data for AFF1108

Batch age (h)	Mean d ($10^{-6}m$)	Median conc. size ($10^{-6}m$)	Median biomass conc. (g/L)
6.3300	110.8921	110.8921	0.9400
7.3300	126.2753	140.8018	1.0886
8.3300	140.8018	170.5731	1.2371
9.4200	170.5731	182.3953	1.3989
10.5000	194.8636	232.9957	1.5593
11.3300	181.5976	255.4656	1.6826
12.5800	182.3954	210.9124	1.8682
13.6700	232.9957	213.0226	2.0301
14.5800	255.4655	197.6400	2.1652
15.5000	210.9125	171.0904	2.3019
16.5000	213.0226	164.6026	2.4504
18.0800	183.9173	145.7964	2.6850
19.0800	205.9857	146.0062	2.9381
20.0800	197.6400	129.1078	3.1912
21.1700	171.0905	127.5378	3.4671
21.9200	173.3749	135.9529	3.6569
31.0800	164.6026	105.3656	5.9754
32.0000	156.3328	102.7412	6.2083
33.0000	145.7964	104.1064	6.4614
34.1700	156.0872	98.5921	6.7576
36.0000	146.0062	80.5319	7.2207
38.1700	141.0902	80.6663	7.7700
40.0800	129.1078	71.0417	7.8173
45.5000	140.1164	60.2757	7.9516
55.1700	127.5377	47.0306	8.1913
79.3300	127.5633	49.5023	8.7900
106.0000	135.9529	41.2655	9.4509
126.0000	137.0940	40.7683	10.4480
150.0000	105.3656	35.0131	11.7950
174.0000	104.6368	45.7076	12.3972
198.0000	102.7412	32.2576	12.9994
204.0000	104.4909	30.1120	13.1500

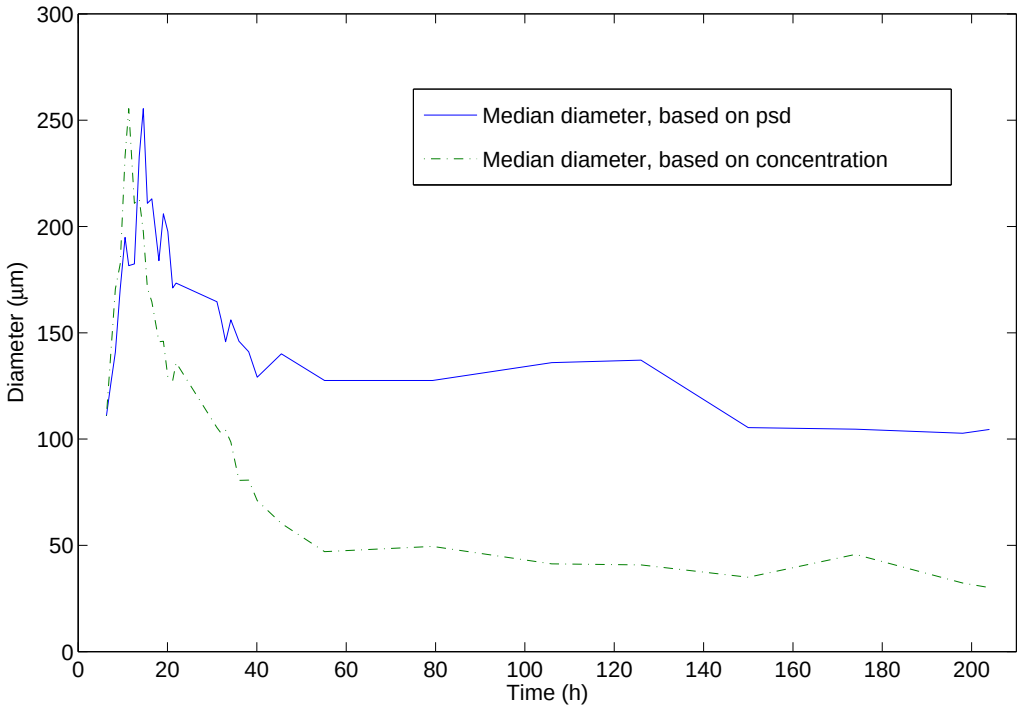


Figure C.4. Median diameters based on both size distribution and concentration/mass distribution

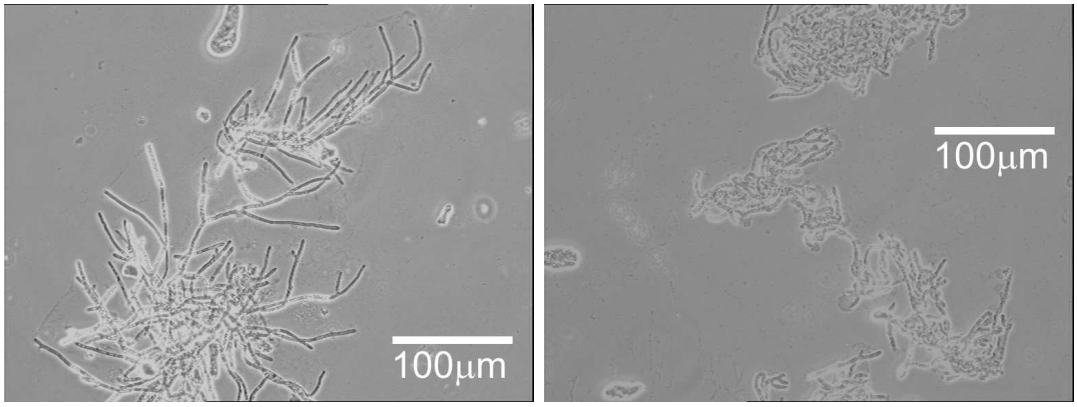
Table C.2. Particle size distribution data for AFF1108

Batch age (h)	Median mass size ($10^{-6}m$)	Median mass (kg)
6.3300	110.8921	1.0961
7.3300	140.8018	1.2619
8.3300	170.5731	1.4364
9.4200	182.3953	1.6203
10.5000	232.9957	1.8046
11.3300	255.4656	1.9472
12.5800	210.9124	2.1548
13.6700	213.0226	2.3395
14.5800	197.6400	2.4911
15.5000	171.0904	2.6438
16.5000	164.6026	2.8239
18.0800	145.7964	3.0891
19.0800	146.0062	3.3574
20.0800	129.1078	3.6467
21.1700	127.5378	3.9518
21.9200	135.9529	4.1576
31.0800	105.3656	6.7180
32.0000	102.7412	6.9678
33.0000	104.1064	7.2456
34.1700	98.5921	7.5580
36.0000	80.5319	8.0690
38.1700	80.6663	8.7431
40.0800	71.0417	8.7660
45.5000	60.2757	9.0016
55.1700	47.0306	9.5909
79.3300	49.5023	11.2478
106.0000	41.2655	13.2312
126.0000	40.7683	15.5706
150.0000	35.0131	18.8147
174.0000	45.7076	21.1956
198.0000	32.2576	23.2854
204.0000	30.1120	23.6062

D

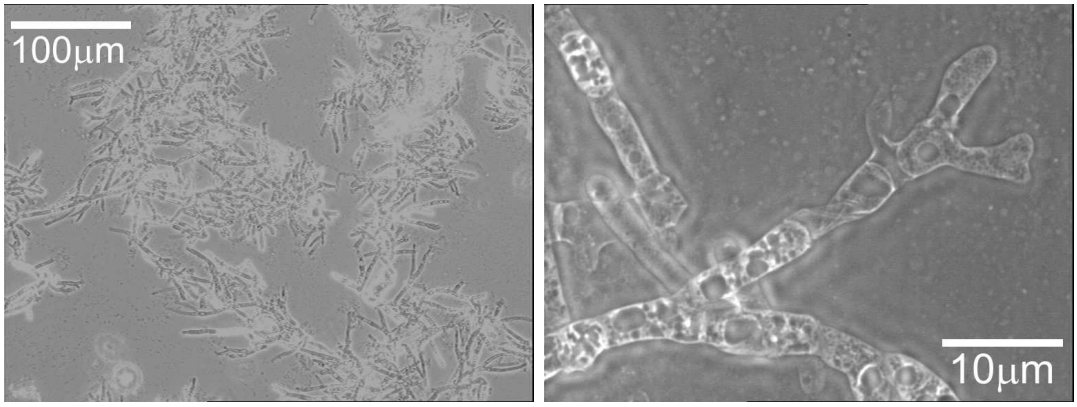
Microscopy images

Images have been taken with a microscope using a 10x, 40x and 100x objective lens. Unless states otherwise the images shown are made with the 10x objective lens. The pictures are given below in chronological order.



(a) 8 hours

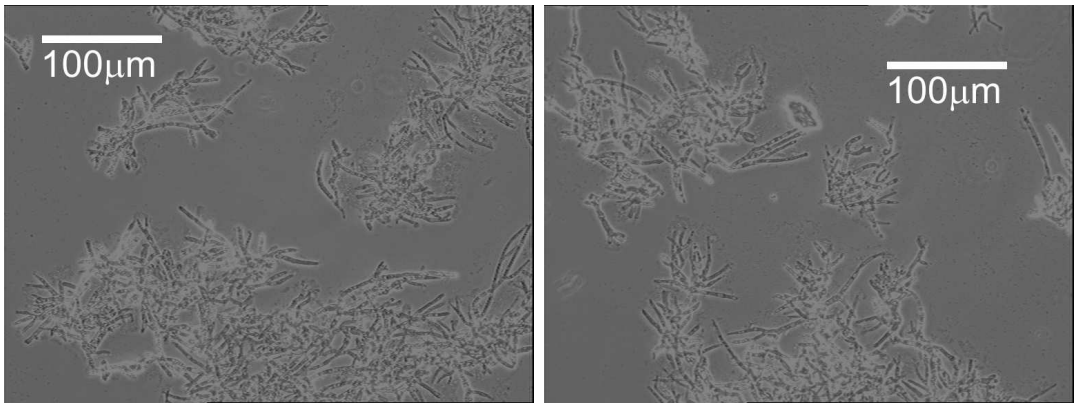
(b) 20 hours



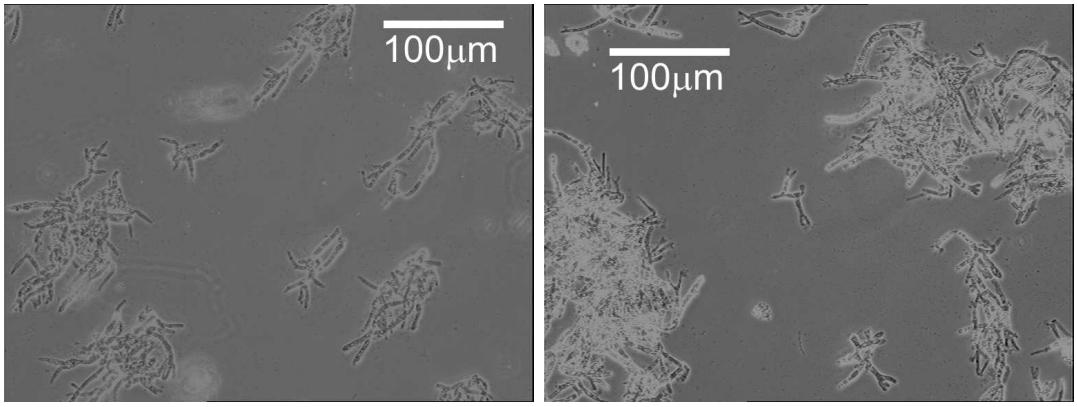
(c) 33 hours

(d) 33 hours - 100x

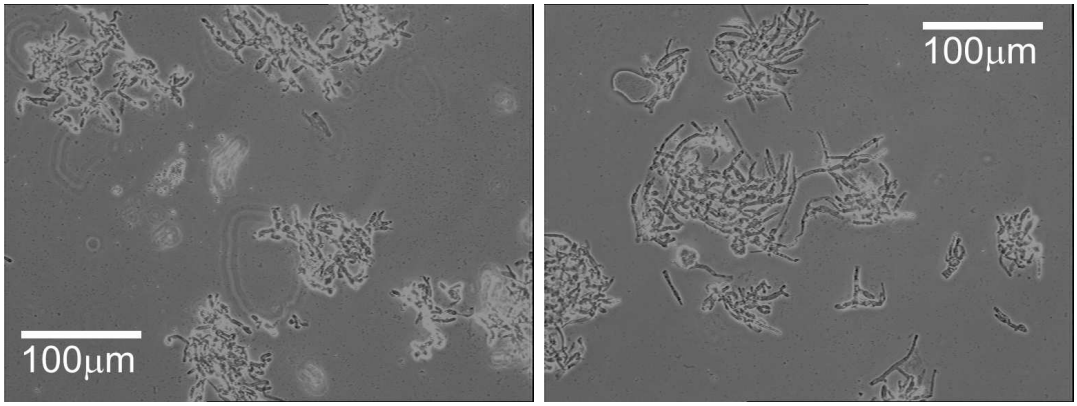
Figure D.1. Images from microscope in chronological order. The images shown are made with a 10x objective lens unless stated otherwise



(a) 35 hours (b) 37 hours

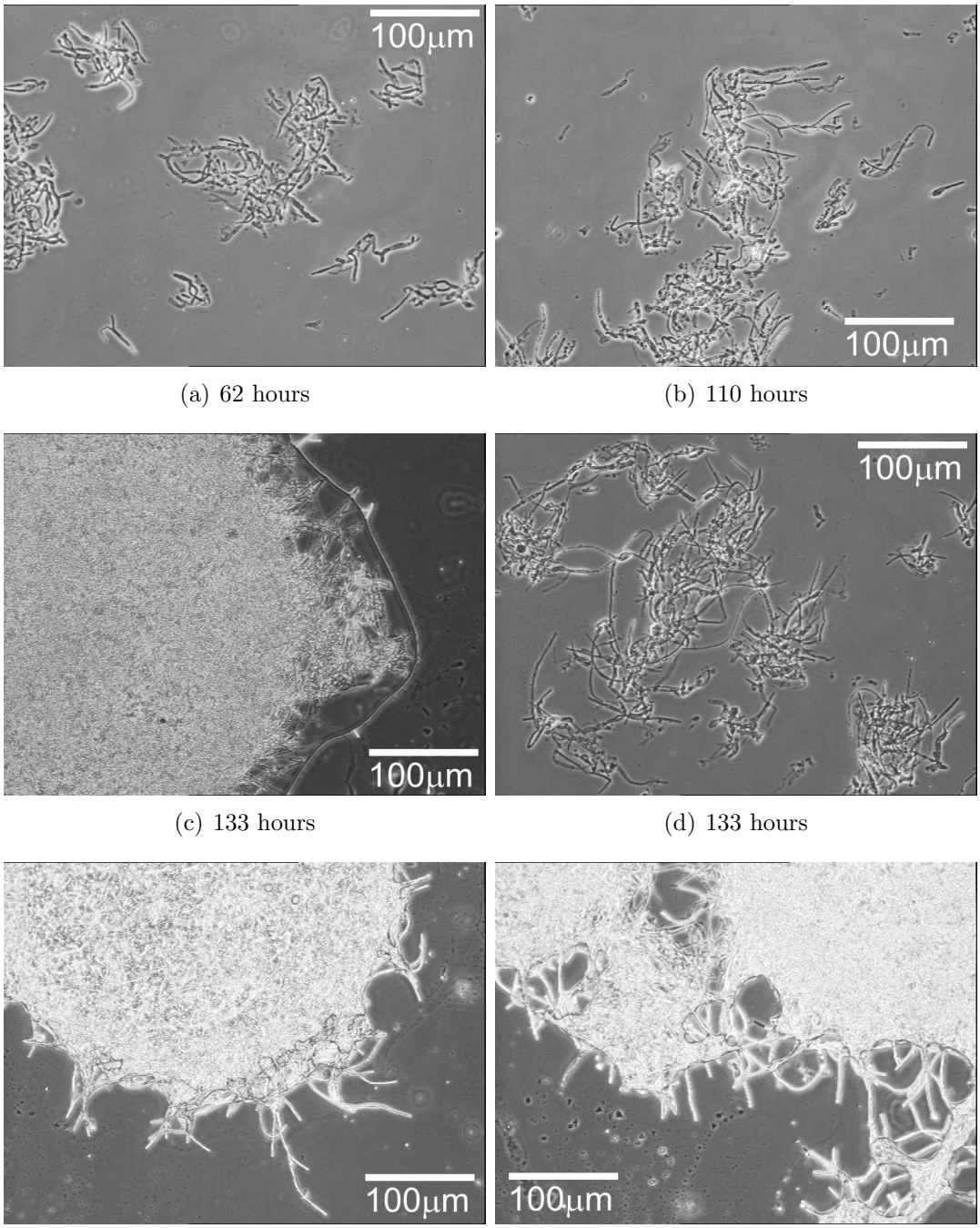


(c) 39 hours (d) 41 hours



(e) 46 hours (f) 55 hours

Figure D.2. Images from microscope in chronological order. The images shown are made with a 10x objective lens unless stated otherwise

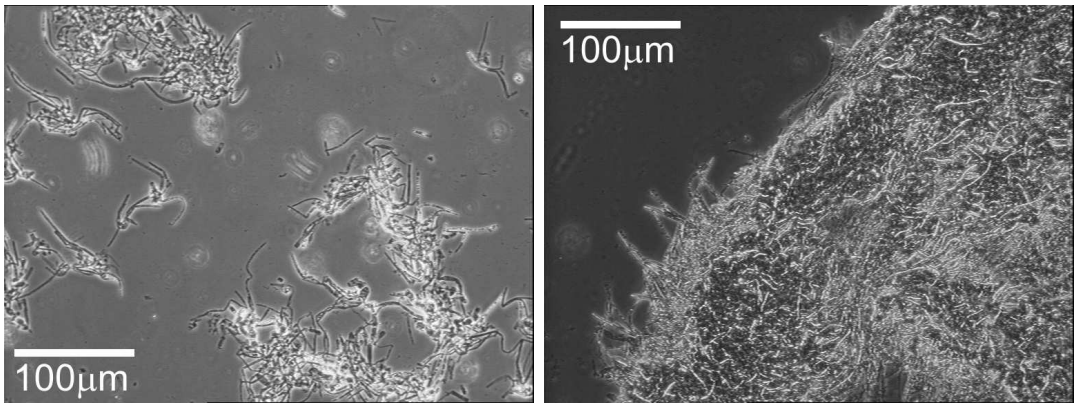


(a) 62 hours (b) 110 hours

(c) 133 hours (d) 133 hours

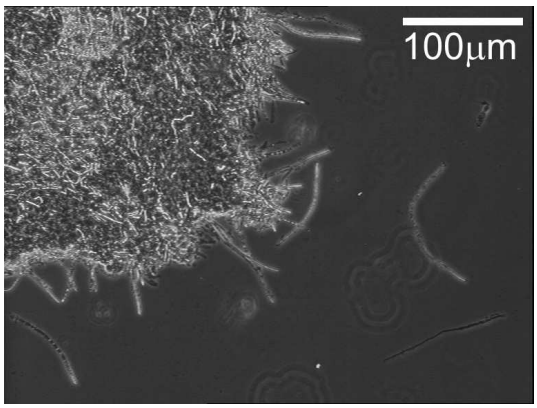
(e) 158 hours (f) 158 hours

Figure D.3. Images from microscope in chronological order. The images shown are made with a 10x objective lens unless stated otherwise

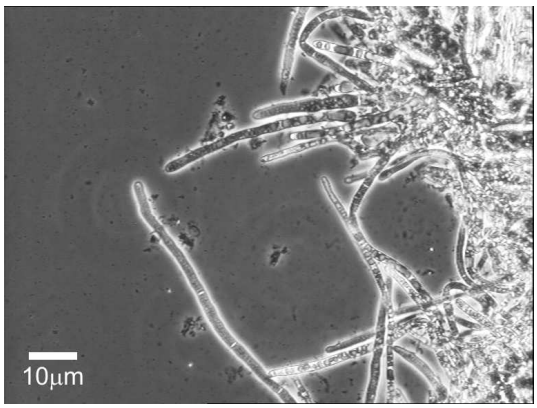


(a) 158 hours

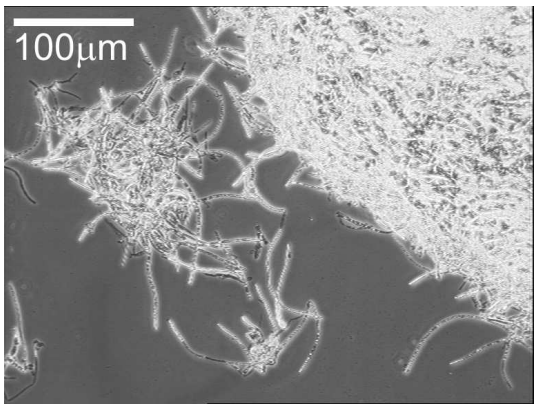
(b) 174 hours



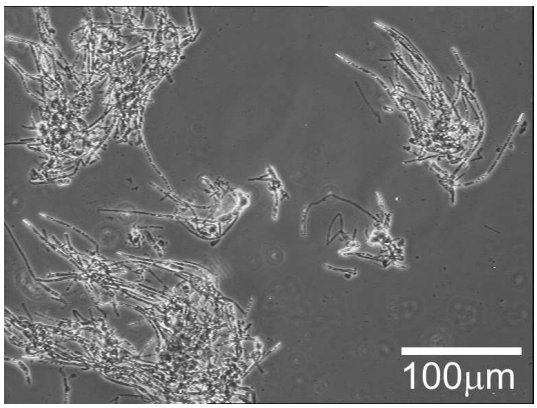
(c) 174 hours



(d) 174 hours - 40x

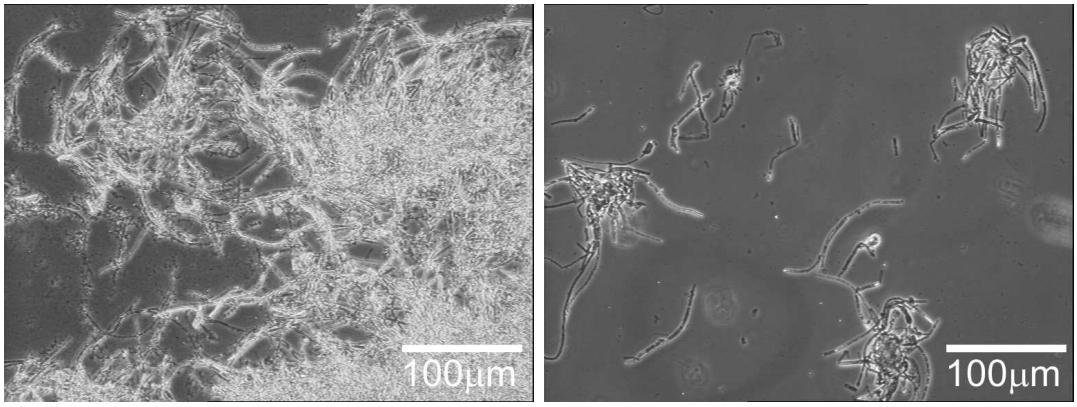


(e) 182 hours

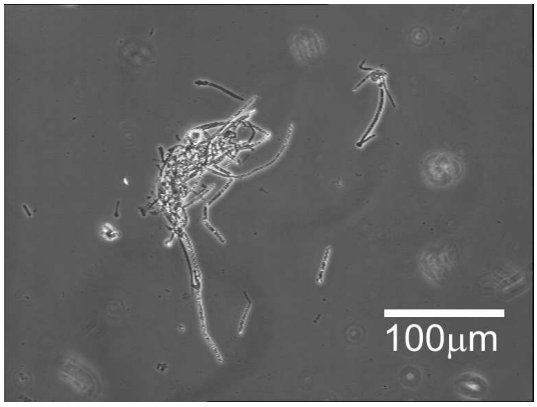


(f) 201 hours

Figure D.4. Images from microscope in chronological order. The images shown are made with a 10x objective lens unless stated otherwise



(a) 203 hours (b) 204 hours



(c) 204 hours

Figure D.5. Images from microscope in chronological order. The images shown are made with a 10x objective lens unless stated otherwise

References

Agger, T. (1999). *Mathematical Modelling of Protein Production in Filamentous Fungi*. Ph.D. thesis, Technical University of Denmark, Denmark.

Agger, T.; Spohr, A. B.; Carlsen, M. and Nielsen, J. (1998). Growth and product formation of Aspergillus oryzae during submerged cultivations: Verification of a morphologically structured model using fluorescent probes. *Biotechnology and Bioengineering*, **57**(3), 321–329.

Bonné, D. (2005). *Optimal and Reproducible Operation of Batch Processes*. Ph.D. thesis, Technical University of Denmark, Denmark.

Bonné, D. and Jørgensen, S. B. (2003). Data-driven modeling of nonlinear and time-varying processes. In *Proceedings to IFAC Symposium on System Identification*, pages 1655–1660.

Bonné, D. and Jørgensen, S. B. (2004). Iterative Learning Model Predictive Control of Batch Processes. *BatchPro Symposium*, pages 67–72. Presented at:.

Brown, D. and Fitzpatrick, S. (1979). A structured model for the kinetics of fungal amylase production. *Biotechnol. Lett.*, **1**, 3–8.

Carlsen, M. (1994). *alpha-Amylase Production by Aspergillus oryzae*. Ph.D. thesis, Technical University of Denmark, Denmark.

Carlsen, M.; Nielsen, J. and Villadsen, J. (1996a). Growth and alpha-amylase production by Aspergillus oryzae during continuous cultivations. *Journal of Biotechnology*, **45**(1), 81–93.

Carlsen, M.; Spohr, A. B.; Nielsen, J. and Villadsen, J. (1996b). Morphology and physiology of an alpha@-amylase producing strain of aspergillus oryzae during batch cultivations. *Biotechnology and Bioengineering*, **49**(3), 266–276.

Gregersen, L. (1999). Supervision of fed-batch fermentations. *Chemical Engineering Journal*, **75**(1), 69–76.

Gregersen, L. (2003). *Monitoring and Fault Diagnosis of Fermentation Processes*. Ph.D. thesis, Technical University of Denmark, Denmark.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering: Theory and Practice using Matlab*. John Wiley &; Sons Inc, New York, NY.

Griffin, D. H. (1994). *Fungal Physiology*. Wiley-Liss, New York.

Kosanovich, K.; Piovoso, M.; Dahl, K.; MacGregor, J. and Nomikos, P. (1994). Multi-Way PCA Applied To An Industrial Batch Process. *American Control Conference, 1994*, **2**, 1294–1298.

Kourti, T.; Nomikos, P. and MacGregor, J. F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal of Process Control*, **5**(4), 277–284.

Kristensen, N. and Madsen, H. (2003). *Continuos Time Stochastic Modelling Mathematics Guide*. Informatics and Mathematical Modelling, Technical University of Denmark.

Kristensen, N.; Madsen, H. and Jorgensen, S. (2004). A method for systematic improvement of stochastic grey-box models. *Computers and Chemical Engineering*, **28**(8), 1431–1449.

Kristensen, N. R. (2003). *Fed-Batch Process Modelling for State Estimation and Optimal Control*. Ph.D. thesis, Technical University of Denmark, Denmark.

Li, Z. J.; Shukla, V.; Fordyce, A. P.; Pedersen, A. G.; Wenger, K. S. and Marten, M. R. (2000). Fungal morphology and fragmentation behavior in a fed-batch Aspergillus oryzae fermentation at the production scale. *Biotechnology and Bioengineering*, **70**(3), 300–312.

Li, Z. J.; Shukla, V.; Wenger, K. S.; Fordyce, A. P.; Pedersen, A. G. and Marten, M. R. (2002). Effects of Increased Impeller Power in a Production-Scale Aspergillusoryzae Fermentation. *Biotechnology Progress*, **18**(3), 437–444.

MacGregor, J. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, **3**(3), 403–414.

Maciejowski, J. M. (2002). *Predictive Control with Constraints*. Pearson Education, Essex, England.

Madigan, M. T.; Martinko, J. M. and Parker, J. (2000). *Brock biology of microorganisms*. Prentice Hall, Upper Saddle River, NJ.

Mayne, D.; Rawlings, J.; Rao, C. and Scokaert, P. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, **36**(6), 789–814.

Megee, R.; Kinoshita, S.; Fredrickson, A. and Tsuchiya, H. (1970). Differentiation and product formation in molds. *Biotechnology and Bioengineering*, **12**, 771–801.

Metz, B. and Kossen, N. (1977). The Growth of Molds in the Form of Pellets - A Literature Review. *Biotechnology and Bioengineering*, **19**, 781–799.

Morari, M. and H. Lee, J. (1999). Model predictive control: past, present and future. *Computers & Chemical Engineering*, **23**(4-5), 667 – 682.

Moreira, M. T.; Sanroman, A.; Feijoo, G. and Lema, J. M. (1996). Control of pellet morphology of filamentous fungi in fluidized bed bioreactors by means of a pulsing flow. Application to Aspergillus niger and Phanerochaete chrysosporium. *Enzyme and Microbial Technology*, **19**(4), 261–266.

Muske, K. R. and Rawlings, J. B. (1993). Model predictive control with linear models. *AIChE Journal*, **39**(2), 262–287.

Nestaas, E. and Wang, D. I. C. (1983). Computer Control of the Penicillin Fermentation using the Filtration Probe in Conjunction with a Structured Process Model. *Biotechnology and Bioengineering*, **25**(3), 781–796.

Nielsen, J. and Villadsen, J. (1992). Modelling of microbial kinetics. *Chemical Engineering Science*, **47**(17), 4225–4270.

Nielsen, J.; Madsen, H. and Young, P. (2000). Parameter estimation in stochastic differential equations: An overview. *Annual Reviews in Control*, **24**, 83–94.

Nielsen, J.; Villadsen, J. and Liden, G. (2003). *Bioreaction Engineering Principles*. Wiley-Liss, New York.

Paul, G. and Thomas, C. (1996). Structured model for hyphal differentiation and penicillin production using Penicillium chrysogenum. *Biotechnology and Bioengineering*, **51**(5), 558–572.

Petersen, N. (2006). *Multivariable Modelling for Control of Industrial Fed-batch Cultivations*. Master's thesis, Technical University of Denmark, Denmark.

Pirt, S. J. (1966). A Theory of the Mode of Growth of Fungi in the Form of Pellets in Submerged Culture. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **166**(1004), 369–373.

Poulsen, B. R. (2005). *Characterization of pathway engineered strains of filamentous fungi in submerged cultures*. Ph.D. thesis, Wageningen University, The Netherlands.

Rao, C. V. and Rawlings, J. B. (2000). Linear programming and model predictive control. *Journal of Process Control*, **10**(2-3), 283–289.

Reichl, U.; King, R. and Gilles, E. (1992). Characterization of pellet morphology during submerged growth of Streptomyces tendae by image analysis. *Biotechnology and Bioengineering*, **39**(2), 164–170.

Spohr, A.; Carlsen, M.; Nielsen, J. and Villadsen, J. (1997). Morphological characterization of recombinant strains of Aspergillus oryzae producing alpha-amylase during batch cultivations. *Biotechnology Letters*, **19**(3), 257–262.

Spohr, A.; Carlsen, M.; Nielsen, J. and Villadsen, J. (1998a). @a-Amylase production in recombinant Aspergillus oryzae during fed-batch and continuous cultivations. *Journal of Fermentation and Bioengineering*, **86**(1), 49–56.

Spohr, A.; Dam-Mikkelsen, C.; Carlsen, M.; Nielsen, J. and Villadsen, J. (1998b). On-line study of fungal morphology during submerged growth in a small flow-through cell. *Biotechnology and Bioengineering*, **58**(5), 541–553.

Stocks, S. (2005). Personal communication.

Wang, N. S. and Stephanopoulos, G. (1983). Application of Macroscopic Balances to the Identification of Gross Measurement Errors. *Biotechnology and Bioengineering*, **25**(9), 2177–2208.

Watson, M. J.; Liakopoulos, A.; Brzakovic, D. and Georgakis, C. (1998). Practical assessment of process data compression techniques. *Industrial &; Engineering Chemistry Research*, **37**(1), 267–274.

Zangirolami, T. C. (1998). *Modeling of Growth and Products Formation in Submerged Cultures of Filamentous Fungi*. Ph.D. thesis, Technical University of Denmark, Denmark.